

FT Protein 400. .575
 /note= "Zeta membrane spanning and intracellular domain"
 XX
 XX MO9215322-A1.
 XX
 XX 17-SEP-1992.
 XX
 XX 06-MAR-1992; 92WO-US001785.
 XX
 XX 07-MAR-1991; 91US-00665961.
 XX
 XX (GEHO) GEN HOSPITAL CORP.
 XX
 XX Seed B, Romeo C, Kolanus W;
 XX
 XX WPI: 1992-331474/40.
 XX
 XX N-PSDB; AAQ28704.
 XX
 XX Therapeutic cells expressing chimeric receptors - directing cellular
 PT response to an infective agent, useful in treating HIV-1, AIDS
 PT Pneumocystis carinii infections etc.
 XX
 XX Example 2; Page 72-73; 114pp; English.
 XX
 XX This sequence represents a fusion protein between the CD4 extracellular
 CC domain and the zeta protein membrane spanning domain and intracellular
 CC domain. Zeta is a 32 kD type I integral membrane homodimer which has a 9
 CC residue extracellular domain and a 112/113 residue intracellular domain
 CC for mouse and human protein respectively. In the production of the CD4
 CC receptor chimera, the zeta cDNA was isolated from the HBB-A17 tumour cell
 CC line and from human natural killer cells. The zeta cDNA was joined to the
 CC extracellular domain of an engineered form of CD4 possessing a BamHI site
 CC just upstream of the membrane spanning domain, by a BamI site naturally
 CC present a few residues upstream of the membrane spanning domain. (Updated
 CC on 25-MAR-2003 to correct FN field.)
 XX
 XX Sequence 575 AA;
 SQ

Query Match 59.7%; Score 2036.5; DB 2; Length 575;
 Best Local Similarity 98.3%; Pred. No. 66-102;
 Matches 397; Conservative 1; Mismatches 3; Indels 3; Gaps 1;

QY 1 MNRGVPRHLLLVQLALPPATQGNKVLGKGGDTVELCTASQKSIQPHMNSNOIK 60
 DB 1 MNRGVPRHLLLVQLALPPATQGNKVLGKGGDTVELCTASQKSIQPHMNSNOIK 60
 QY 61 ILGNQGSFLTKGSPSKLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEVOL 120
 DB 61 ILGNQGSFLTKGSPSKLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEVOL 120
 QY 121 LVFGLTANSPDTHLLQGOSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDG 180
 DB 121 LVFGLTANSPDTHLLQGOSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDG 180
 QY 181 TWCTCTVQONQKVEFKIDIVLAFQKASSIYKKEGEOVESFPLATFVEKLTGSGELMW 240
 DB 181 TWCTCTVQONQKVEFKIDIVLAFQKASSIYKKEGEOVESFPLATFVEKLTGSGELMW 240
 QY 241 QABRASSSSKSWITFDLKNKEVSVKRVYTOPDKLQMGKKLPLHLTLPOALPOYAGSGNITLA 300
 DB 241 QABRASSSSKSWITFDLKNKEVSVKRVYTOPDKLQMGKKLPLHLTLPOALPOYAGSGNITLA 300
 QY 301 LEATGTGLHGEVNLVVMRATQLOKNTLCEVWGPTSPKLMSTLKENKEAKVSKREKRVW 360
 DB 301 LEATGTGLHGEVNLVVMRATQLOKNTLCEVWGPTSPKLMSTLKENKEAKVSKREKRVW 360
 QY 361 LNPEAGMMQCLLSDSGOVLLESNIKVLPTWSTPVPAPAPKSC 404
 DB 361 LNPEAGMMQCLLSDSGOVLLESNIKVLPTWSTPVPAPAPKSC 404

ID AAR78676 standard; protein; 575 AA.
 XX
 XX AAR78676;
 AC
 XX 16-APR-1996 (first entry)
 DT
 XX
 XX T-cell receptor zeta.
 DE
 XX
 XX Chimeric receptor; CD4; T-cell receptor zeta; HIV; cytolysis;
 KM human immunodeficiency virus; adoptive immunotherapy.
 XX
 XX Homo sapiens.
 OS
 XX
 XX MO9521528-A1.
 XX
 XX 17-AUG-1995.
 XX
 XX 12-JAN-1995; 95WO-US000454.
 XX
 XX 14-FEB-1994; 94US-00195395.
 PR 02-AUG-1994; 94US-00284391.
 XX
 XX (GEHO) GEN HOSPITAL CORP.
 XX
 XX Seed B, Banapour B, Romeo C, Kolanus W;
 XX
 XX WPI: 1995-292893/38.
 XX
 XX N-PSDB; AAQ96122.
 DR
 XX Target cytolysis of HIV-infected cells - by chimeric CD4 receptor-bearing
 PT cells.
 PT
 XX
 XX Example 2; Page 76-77; 118pp; English.
 PS
 XX Fusion proteins comprising the extracellular domain of CD4 fused to T-
 CC cell receptor zeta, gamma or eta (AAR78676-78, respectively) were
 CC expressed in CV1 using a vaccine virus vector. These CD4:zeta, CD4:gamma
 CC and CD4:eta chimeric receptors mediated cytolysis of targets expressing
 CC HIV gp120/41
 XX
 XX Sequence 575 AA;
 SQ

Query Match 59.7%; Score 2036.5; DB 2; Length 575;
 Best Local Similarity 98.3%; Pred. No. 66-102;
 Matches 397; Conservative 1; Mismatches 3; Indels 3; Gaps 1;

QY 1 MNRGVPRHLLLVQLALPPATQGNKVLGKGGDTVELCTASQKSIQPHMNSNOIK 60
 DB 1 MNRGVPRHLLLVQLALPPATQGNKVLGKGGDTVELCTASQKSIQPHMNSNOIK 60
 QY 61 ILGNQGSFLTKGSPSKLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEVOL 120
 DB 61 ILGNQGSFLTKGSPSKLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEVOL 120
 QY 121 LVFGLTANSPDTHLLQGOSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDG 180
 DB 121 LVFGLTANSPDTHLLQGOSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDG 180
 QY 181 TWCTCTVQONQKVEFKIDIVLAFQKASSIYKKEGEOVESFPLATFVEKLTGSGELMW 240
 DB 181 TWCTCTVQONQKVEFKIDIVLAFQKASSIYKKEGEOVESFPLATFVEKLTGSGELMW 240
 QY 241 QABRASSSSKSWITFDLKNKEVSVKRVYTOPDKLQMGKKLPLHLTLPOALPOYAGSGNITLA 300
 DB 241 QABRASSSSKSWITFDLKNKEVSVKRVYTOPDKLQMGKKLPLHLTLPOALPOYAGSGNITLA 300
 QY 301 LEATGTGLHGEVNLVVMRATQLOKNTLCEVWGPTSPKLMSTLKENKEAKVSKREKRVW 360
 DB 301 LEATGTGLHGEVNLVVMRATQLOKNTLCEVWGPTSPKLMSTLKENKEAKVSKREKRVW 360
 QY 361 LNPEAGMMQCLLSDSGOVLLESNIKVLPTWSTPVPAPAPKSC 404
 DB 361 LNPEAGMMQCLLSDSGOVLLESNIKVLPTWSTPVPAPAPKSC 404

```

RESULT 1
AAR89456
ID AAR89456 standard; protein; 575 AA.
XX
XX AAR89456;
XX
XX DT 26-SEP-1996 (first entry)
XX
XX CD4;zeta fusion protein.
XX
XX KW CD7; transmembrane domain; chimeric receptor; CD5; CD34; CH2; CH3; IgG1;
KW human; CD4; HIV; proteinaceous alpha-helix; T cell; B cell; neutrophil;
KW dendritic cell; therapy; mammal; infection.
XX
XX OS Synthetic.
XX
XX PN WO9603883-A1.
XX
XX PD 15-FEB-1996.
XX
XX PE 26-JUL-1995; 95MO-US009468.
XX
XX PR 02-AUG-1994; 94US-00284391.
XX
XX PR 24-FEB-1995; 95US-00394388.
XX
XX PA (GENO ) GEN HOSPITAL CORP.
XX
XX PI Seed B, Banapour B, Romeo C, Kolanus W;
XX
XX DR WPI; 1996-129034/13.
XX
XX DR N-PSDB; AAT10801.
XX
XX PT Membrane-bound chimeric receptor comprising extracellular portion
PT including CD4 fragment - cells expressing receptor can be used for
PT treatment of HIV infection.
XX
XX PS Example 2; Page 77-78; 134pp; English.
XX
CC AAT10801-T10803 represent membrane bound proteinaceous chimeric receptors
CC of the invention. This sequence represents the CD4;zeta chimera. The
CC transmembrane region of the chimeric receptor acts to separate the
CC intracellular and extracellular domains of the chimera, and contains a
CC portion of the CD7 (see AAR89440), CD5 or CD34 transmembrane domains.
CC Alternatively, the extracellular portion of the receptor can be separated
CC from the intracellular domain by the hinge, CH2 and CH3 domains of human
CC IgG1 (see AAR89441). The extracellular portion of the chimeric receptor
CC contains a fragment of CD4 (amino acids 1-394 or 1-200 of the CD4
CC sequence, see AAR89450 and AAR89451) which specifically recognises and
CC binds HIV-infected cells, but does not mediate HIV infection. The
CC extracellular domain of the receptor is separated from the cell membrane
CC by 48 or 72 angstroms, or by one or more proteinaceous alpha-helices. The
CC cells expressing the receptor are preferably T cells, B cells,
CC neutrophils, or dendritic cells. The therapeutic cells expressing the
CC chimeric receptor are administered to a mammal to treat HIV infection
CC
XX
SQ Sequence 575 AA;
XX
Query Match 59.7%; Score 2036.5; DB 2; Length 575;
Best Local Similarity 98.3%; Pred. No. 6e-102;
Matches 397; Conservative 1; Mismatches 3; Indels 3; Gaps 1
QY 1 MNRGVDFRHLVLLQLALLPAATQGKVVILGGKGDTVELTCTASOKKSIOFHKNSNQIK 60
Db 1 MNRGVDFRHLVLLQLALLPAATQGKVVILGGKGDTVELTCTASOKKSIOFHKNSNQIK 60
QY 61 ILGNQGSFLTCKPGSKLNDRAISRSLMDQGNFLIIKNLTIESDPYICVEVOKEEVL 120
Db 61 ILGNQGSFLTCKPGSKLNDRAISRSLMDQGNFLIIKNLTIESDPYICVEVOKEEVL 120
QY 121 LVFGTLTANSDTHLLOQSITLTLESBPGSSPVQCRRPRGKNIQGGKTISVSQLELDGSG 180

```

Db	121	LVFGITANSDBTHLLOQSGSLTTLESPPSSPSVQCRSPGKNITQSGKTLVSQLEIQDSG	180
Qy	181	TWTCVTLONQKVKFEKIDIVLAFQASSIVYKKEGEVSEFPLAFTVEKLTGSGELMW	240
Db	181	TWTCVTLONQKVKFEKIDIVLAFQASSIVYKKEGEVSEFPLAFTVEKLTGSGELMW	240
Qy	241	QAEKSSSKSWITFDPLKNKEVSVKRVTDPPKLOMGKULPLHLTLPQALPQYAGSGNLTIA	300
Db	241	QAEKSSSKSWITFDPLKNKEVSVKRVTDPPKLOMGKULPLHLTLPQALPQYAGSGNLTIA	300
Qy	301	LEAKTGKLGHOVNLVYMBATOLQKLTCEWNGPSPKMLSLKLENKAKVSKREKPVWV	360
Db	301	LEAKTGKLGHOVNLVYMBATOLQKLTCEWNGPSPKMLSLKLENKAKVSKREKPVWV	360
Qy	361	LNPEAGMWQCLSDSGQVLTLESNIKVLPTWSTPVPCEPAPKPKSC	404
Db	361	LNPEAGMWQCLSDSGQVLTLESNIKVLPTWSTPVPCEPAPKPKSC	404
RESULT 42			
AAW02213			
ID	AAW02213	standard; protein; 575 AA.	
XX	AAW02213;		
XX	11-NOV-1996	(first entry)	
DE	CD4:T-cell receptor zeta chain chimaeric receptor.		
XX	Chimaeric receptor; cellular immunity; adoptive immunotherapy; CD4;		
KM	human immunodeficiency virus type 1; HIV-1; AIDS; therapy;		
KW	T-cell receptor zeta chain; cytotoxic T lymphocyte; CTL.		
XX	Os sapiens.		
XX			
PH	Key	Location/Qualifiers	
FT	Domain	1..393	
FT		/label= "Extracellular domain	
FT		/note= "CD4 extracellular domain"	
FT	Region	394..396	
FT		/label= "Linker	
FT		/note= "encoding DNA contains a BamHI site used for	
FT		fusion construction"	
FT	Region	397..575	
FT		/note= "Region of fusion derived from zeta chain,	
FT		preferred signal-transducing portions for constructs of	
FT		the invention are amino acids 421-575, 423-255, 438-455,	
FT		461-494, 494-528, 400-420 and 421-462"	
FT	Domain	400..437	
FT		/label= "transmembrane domain	
FT		/note= "zeta chain transmembrane domain"	
FT	Domain	438..575	
FT		/label= "intracellular domain	
FT		/note= "zeta chain intracellular domain"	
XX			
PN	WO9625953-A1.		
XX			
XX	29-AUG-1996.		
XX			
XX	25-JAN-1996;	96WO-US001056.	
XX			
PR	24-FEB-1995;	95US-00394176.	
XX			
PA	(GHEO) GEN HOSPITAL CORP.		
PI	Seed B, Romeo C, Kolanus W;		
XX			
DR	WPI; 1996-402134/40.		
DR	N-PSDB; AAT36758.		
PT	Direction of cellular immune response using therapeutic cell expressing 2		
PT	chimaeric receptors - comprising region binding to target cell and region		
PT	that signals target cell destruction, or CD28 region, partic. for		

PT eliminating HIV-infected cells.
XX
XX Claim 7; Page 74-75; 120pp; English.
XX
CC A chimeric receptor (AAW00213) comprises the extracellular domain of an
PS engineered form of the CD4 cellular receptor for HIV and the
XX transmembrane and intracellular regions, including the cytoytic signal-
CC transducing portion, of the human T-cell receptor zeta chain; the region
CC of the fusion is shown in AAW02221. It can be obid. by inserting a gene
CC fusion (AAT36758) into a vaccinia virus vector and expression in host
CC cells. Chimeric receptors comprising CD4 fused to zeta, eta (see also
CC AAW02215) or Fc receptor gamma (see also AAW02214) chains are capable of
CC directing cytotoxic T lymphocytes to specifically recognise and kill
CC cells expressing HIV gp120, thus providing a therapy for AIDS
XX
SQ Sequence 575 AA;
Query Match 59.7%; Score 2036.5; DB 2; Length 575;
Best Local Similarity 98.3%; Pred. No. 6e-102;
Matches 397; Conservative 1; Mismatches 3; Indels 3; Gaps 1;
QY 1 MNRGVPRHLLLVQLALPAAIQGNKRVLGKGDVETLCTASQKKSIOFHMKNNSQIK 60
DB 1 MNRGVPRHLLLVQLALPAAIQGNKRVLGKGDVETLCTASQKKSIOFHMKNNSQIK 60
QY ILGNQGSFLTKGPSKLNDRADSRSLMDQGNFPLIINLKI ESDTYICEVEDQKEEVOL 120
DB 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNFPLIINLKI ESDTYICEVEDQKEEVOL 120
QY 121 LVFGLTANSDTHLLQGGSLTLTLSPGSSPSVOCSPRGKNIQGGKTLVSQLELDDSG 180
DB 121 LVFGLTANSDTHLLQGGSLTLTLSPGSSPSVOCSPRGKNIQGGKTLVSQLELDDSG 180
QY 181 TWCTCTVLOKQKVEFKIDIVLAFOKASSIYKKEGEVSEFPLAFVETKLTSGGELMW 240
DB 181 TWCTCTVLOKQKVEFKIDIVLAFOKASSIYKKEGEVSEFPLAFVETKLTSGGELMW 240
QY 241 QAEBASSSKSWITFDLNKKEVSVRKVTQDPKLOMGKKLPLHLTLPOALPOYAGSGNLTIA 300
DB 241 QAEBASSSKSWITFDLNKKEVSVRKVTQDPKLOMGKKLPLHLTLPOALPOYAGSGNLTIA 300
QY 301 LEAKTGKLGHEVNLVYWRATOLQKNTLCEVWGPTSPKLMSTLLENKEAVSKREKPVW 360
DB 301 LEAKTGKLGHEVNLVYWRATOLQKNTLCEVWGPTSPKLMSTLLENKEAVSKREKPVW 360
QY 361 LNPEAGMOCILSDSGVLLSNIKVLPTWSTPVHAA--DPKLC 401
DB 361 LNPEAGMOCILSDSGVLLSNIKVLPTWSTPVHAA--DPKLC 401
RESULT 43
AAW83140
ID AAW83140 standard; protein; 575 AA.
XX
XX AAW83140;
XX
XX 03-FEB-1999 (first entry)
XX
DE Chimeric receptor containing human zeta polypeptide.
XX
XX Human; zeta; eta; gamma; membrane-bound chimeric receptor; infection;
KW tumour; cancer cell; autoimmune-generated cell; T cell receptor; CD3;
KW CD4; B cell receptor; Fc receptor; pathogen; bacterial; fungal;
KW protozoan; viral.
XX
XX Synthetic.
OS Homo sapiens.
XX
XX US5843728-A.
XX
XX 01-DEC-1998.
XX
XX 05-APR-1995; 95US-00417495.

XX
XX 07-MAR-1991; 91US-00665961.
PR 06-MAR-1992; 92US-00847566.
PR 28-FEB-1994; 94US-00203866.
XX
XX (GEHO) GEN HOSPITAL CORP.
XX
XX Romeo C, Kolanus W, Seed B;
XX
XX WPI; 1999-044582/04.
XX
XX N-PSDB; AAV70156.
XX
XX Membrane-bound chimeric receptors - comprising extracellular portion
PT which recognises and binds a target cell and an intracellular portion of
PT e.g. a T-cell receptor.
XX
XX Example 2; Col 39-42; 57pp; English.
XX
XX The present invention describes DNA encoding a membrane-bound chimeric
CC receptor comprising: (a) an extracellular portion that specifically
CC recognises and binds a target cell or a target infective agent; and (b)
CC an intracellular portion of a T-cell receptor CD3, zeta or eta
CC polypeptide, a B-cell receptor polypeptide or an Fc receptor polypeptide.
CC The present sequence represents a chimeric receptor containing the human
CC zeta polypeptide. Cells expressing chimeric receptors of the present
CC invention can be administered to mammals in order to destroy pathogens
CC (e.g. bacteria, fungi, protozoa or viruses, especially HIV), cancer cells
CC or autoimmune-generated cells
XX
SQ Sequence 575 AA;
Query Match 59.7%; Score 2036.5; DB 2; Length 575;
Best Local Similarity 98.3%; Pred. No. 6e-102;
Matches 397; Conservative 1; Mismatches 3; Indels 3; Gaps 1;
QY 1 MNRGVPRHLLLVQLALPAAIQGNKRVLGKGDVETLCTASQKKSIOFHMKNNSQIK 60
DB 1 MNRGVPRHLLLVQLALPAAIQGNKRVLGKGDVETLCTASQKKSIOFHMKNNSQIK 60
QY ILGNQGSFLTKGPSKLNDRADSRSLMDQGNFPLIINLKI ESDTYICEVEDQKEEVOL 120
DB 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNFPLIINLKI ESDTYICEVEDQKEEVOL 120
QY 121 LVFGLTANSDTHLLQGGSLTLTLSPGSSPSVOCSPRGKNIQGGKTLVSQLELDDSG 180
DB 121 LVFGLTANSDTHLLQGGSLTLTLSPGSSPSVOCSPRGKNIQGGKTLVSQLELDDSG 180
QY 181 TWCTCTVLOKQKVEFKIDIVLAFOKASSIYKKEGEVSEFPLAFVETKLTSGGELMW 240
DB 181 TWCTCTVLOKQKVEFKIDIVLAFOKASSIYKKEGEVSEFPLAFVETKLTSGGELMW 240
QY 241 QAEBASSSKSWITFDLNKKEVSVRKVTQDPKLOMGKKLPLHLTLPOALPOYAGSGNLTIA 300
DB 241 QAEBASSSKSWITFDLNKKEVSVRKVTQDPKLOMGKKLPLHLTLPOALPOYAGSGNLTIA 300
QY 301 LEAKTGKLGHEVNLVYWRATOLQKNTLCEVWGPTSPKLMSTLLENKEAVSKREKPVW 360
DB 301 LEAKTGKLGHEVNLVYWRATOLQKNTLCEVWGPTSPKLMSTLLENKEAVSKREKPVW 360
QY 361 LNPEAGMOCILSDSGVLLSNIKVLPTWSTPVHAA--DPKLC 401
DB 361 LNPEAGMOCILSDSGVLLSNIKVLPTWSTPVHAA--DPKLC 401
RESULT 44
AAR27277
ID AAR27277 standard; protein; 462 AA.
XX
XX AAR27277;
XX
XX 25-MAR-2003 (revised)
XX
XX 28-JUL-1995 (first entry)
XX

```

DE CD4:eta peptide chimeric protein.
XX
XX Fusion protein; CD4; extracellular domain; zeta; eta; gamma;
KW membrane spanning domain; intracellular domain; type I;
KW integral membrane homodimer; TCR; T cell antigen receptor;
KW extracellular domain; mouse; human; receptor; chimera;
KW HPB-ALL tumour cell line; natural killer cell.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
XX FH 1..399
XX FT Protein /note="CD4 extracellular domain"
XX FT 400..462
XX FT Protein /note="Zeta membrane spanning and intracellular domain"
XX
XX MO9215322-A1.
XX
XX PD 17-SEP-1992.
XX
XX PF 06-MAR-1992; 92WO-US001785.
XX
XX PR 07-MAR-1991; 91US-00665961.
XX
XX PA (GENO) GEN HOSPITAL CORP.
XX
XX PI Seed B, Romeo C, Kolanus W;
XX
XX DR WPI; 1992-331474/40.
XX
XX DR N-PSDB; AAQ28705.
XX
XX PT Therapeutic cells expressing chimeric receptors - directing cellular
XX response to an infective agent, useful in treating HIV-1, AIDS
XX Pneumocystis carinii infections etc.
XX
XX Example 2; Page 73-74; 114pp; English.
XX
XX PS This sequence represents a fusion protein between the CD4 extracellular
XX domain and the eta protein membrane spanning domain and intracellular
XX domain. Eta is an isoform of zeta (see also AAR27276) which is a 32 kD
XX type I integral membrane homodimer, which arises by alternate RNA
XX splicing. It is present in reduced amounts in cells expressing the T cell
XX antigen receptor. Zeta-eta heterodimers are thought to mediate the
XX formation of inositol phosphates, as well as the receptor initiated cell
XX death called apoptosis. In the production of the CD4 receptor chimera,
XX the eta cDNA was isolated from the HPB-ALL tumour cell line and from
XX human natural killer cells. The eta cDNA was joined to the extracellular
XX domain of an engineered form of CD4 possessing a BamHI site just upstream
XX of the membrane spanning domain, by a BamHI site naturally present a few
XX residues upstream of the membrane spanning domain. (Updated on 25-MAR-
XX 2003 to correct PN field.)
XX
XX SQ Sequence 462 AA;
XX
XX Query Match 59.5%; Score 2032.5; DB 2; Length 462;
XX Best Local Similarity 98.0%; Pred. NO. 7.8e-102;
XX Matches 396; Conservative 2; Mismatches 3; Indels 3; Gaps 1;
XX
XX QY 1 MNRGVPRHLLLVQLALPAATQGNKVVLGKKGDTVELTCTASQKKSIOFHKNSNOIK 60
XX DB 1 MNRGVPRHLLLVQLALPAATQGNKVVLGKKGDTVELTCTASQKKSIOFHKNSNOIK 60
XX
XX QY 61 ILNGQSFLLTKGPKSKLNDRAISRSLMDQGNFPLIINKLIEDSDTYICEVEDQKEVOL 120
XX DB 61 ILNGQSFLLTKGPKSKLNDRAISRSLMDQGNFPLIINKLIEDSDTYICEVEDQKEVOL 120
XX
XX QY 121 LVFGLTANSDTHLLQGOSLTLTLSPGSSPSVQCRSPRGKNIQGKTLISVSQLLEIDSG 180
XX DB 121 LVFGLTANSDTHLLQGOSLTLTLSPGSSPSVQCRSPRGKNIQGKTLISVSQLLEIDSG 180
XX
XX QY 181 TWCTGVANQKQVFFKIDIVLAFQKASSIYKKKGQVFFSPFLATVEKLTGSGELMW 240
XX DB 181 TWCTGVANQKQVFFKIDIVLAFQKASSIYKKKGQVFFSPFLATVEKLTGSGELMW 240
XX
XX 181 TWCTGVANQKQVFFKIDIVLAFQKASSIYKKKGQVFFSPFLATVEKLTGSGELMW 240

```

```

QY 241 QAEARSSKSWITTEPDLKKEVSVKRVTDPPKIQMGKXLPILHILTPQALPQVAGSGNLTLA 300
DB 241 QAEARSSKSWITTEPDLKKEVSVKRVTDPPKIQMGKXLPILHILTPQALPQVAGSGNLTLA 300
QY 301 LEAKTGKTLHQBENLVVMBATQLOKNIJCEVWGPTSPKMLSLKENKEAKVSKREKPVWV 360
DB 301 LEAKTGKTLHQBENLVVMBATQLOKNIJCEVWGPTSPKMLSLKENKEAKVSKREKPVWV 360
QY 361 LNPEAGMWOCILSDSGOVLLESNIVLPTWSTPVPCEPAPKSC 404
DB 361 LNPEAGMWOCILSDSGOVLLESNIVLPTWSTPVPCEPAPKSC 404
XX
XX RESULT 45
XX ID AAR78677 standard; protein; 462 AA.
XX
XX AC AAR78677;
XX
XX DT 16-APR-1996 (first entry)
XX
XX DE T-cell receptor gamma.
XX
XX KW Chimeric receptor; CD4; T-cell receptor gamma; HIV; cytolysis;
XX human immunodeficiency virus; adoptive immunotherapy.
XX
XX OS Homo sapiens.
XX
XX PN MO9521528-A1.
XX
XX PD 17-AUG-1995.
XX
XX PF 12-JAN-1995; 95WO-US000454.
XX
XX PR 14-FEB-1994; 94US-00195395.
XX
XX PR 02-AUG-1994; 94US-00284391.
XX
XX PA (GENO) GEN HOSPITAL CORP.
XX
XX PI Seed B, Banapour B, Romeo C, Kolanus W;
XX
XX DR WPI; 1995-292893/38.
XX
XX DR P-PSDB; AAQ6123.
XX
XX PT Target cytolysis of HIV-infected cells - by chimeric CD4 receptor-bearing
XX cells.
XX
XX PS Example 2; Page 77-78; 118pp; English.
XX
XX XX Fusion proteins comprising the extracellular domain of CD4 fused to T-
XX cell receptor zeta, gamma or eta (AAR78676-78, respectively) were
XX expressed in CV1 using a vaccine virus vector. These CD4:zeta, CD4:gamma
XX and CD4:eta chimeric receptors mediated cytolysis of targets expressing
XX HIV gp120/41
XX
XX SQ Sequence 462 AA;
XX
XX Query Match 59.5%; Score 2032.5; DB 2; Length 462;
XX Best Local Similarity 98.0%; Pred. NO. 7.8e-102;
XX Matches 396; Conservative 2; Mismatches 3; Indels 3; Gaps 1;
XX
XX QY 1 MNRGVPRHLLLVQLALPAATQGNKVVLGKKGDTVELTCTASQKKSIOFHKNSNOIK 60
XX DB 1 MNRGVPRHLLLVQLALPAATQGNKVVLGKKGDTVELTCTASQKKSIOFHKNSNOIK 60
XX
XX QY 61 ILNGQSFLLTKGPKSKLNDRAISRSLMDQGNFPLIINKLIEDSDTYICEVEDQKEVOL 120
XX DB 61 ILNGQSFLLTKGPKSKLNDRAISRSLMDQGNFPLIINKLIEDSDTYICEVEDQKEVOL 120
XX
XX QY 121 LVFGLTANSDTHLLQGOSLTLTLSPGSSPSVQCRSPRGKNIQGKTLISVSQLLEIDSG 180
XX DB 121 LVFGLTANSDTHLLQGOSLTLTLSPGSSPSVQCRSPRGKNIQGKTLISVSQLLEIDSG 180
XX
XX 121 LVFGLTANSDTHLLQGOSLTLTLSPGSSPSVQCRSPRGKNIQGKTLISVSQLLEIDSG 180

```

```

QY 181 TWCTVLONOKKVEFKIDIVLAFQKASSIYKKEGQVEFSFPLAFTVEKLTGSGELMW 240
D 181 TWCTVLONOKKVEFKIDIVLAFQKASSIYKKEGQVEFSFPLAFTVEKLTGSGELMW 240
QY 241 QAERASSSSKSWITFDLKNKEVSVRVTQDPKLGKGLPLHLTLPQALPOYAGSGNLTLA 300
D 241 QAERASSSSKSWITFDLKNKEVSVRVTQDPKLGKGLPLHLTLPQALPOYAGSGNLTLA 300
QY 301 LEATGTLHQBVLVWVRATQLOKNTLCEVWGPTSPKLMSTLKENKEAKVSKREKPVWV 360
D 301 LEATGTLHQBVLVWVRATQLOKNTLCEVWGPTSPKLMSTLKENKEAKVSKREKPVWV 360
QY 361 LNPEAGMOCILSDSGOVLLESNIKVLPTWSTPVPCEAPBPKSC 404
D 361 LNPEAGMOCILSDSGOVLLESNIKVLPTWSTPVPCEAPBPKSC 404

RESULT 46
AAR89457
ID AAR89457 standard; protein; 462 AA.
AC AAR89457;
DT 26-FEB-1996 (first entry)
DE CD4:gamma fusion protein.
KM CD7; transmembrane domain; chimeric receptor; CD5; CD34; CH2; CH3; IgG1;
human; CD4; HIV; proteinaceous alpha-helix; T cell; B cell; neutrophil;
dendritic cell; therapy; mammal; infection.
OS Synthetic.
PN WO9603883-A1.
PD 15-FEB-1996.
PF 26-JUL-1995; 95WO-US009468.
PR 02-AUG-1994; 94US-00284391.
PR 24-FEB-1995; 95US-00394388.
PA (GENO) GEN HOSPITAL CORP.
PI Seed B, Banapour B, Romeo C, Kolanus W;
DR MPI; 1996-129034/13.
DR N-PSDB; AAT10802.
PT Membrane-bound chimeric receptor comprising extracellular portion
including CD4 fragment - cells expressing receptor can be used for
treatment of HIV infection.
PS Example 2; Page 79; 134pp; English.
XX AAT10801-T10803 represent membrane bound proteinaceous chimeric receptors
of the invention. This sequence represents the CD4:gamma chimera. The
transmembrane region of the chimeric receptor acts to separate the
intracellular and extracellular domains of the chimera, and contains a
portion of the CD7 (see AAR89440), CD5 or CD34 transmembrane domains.
Alternatively, the extracellular portion of the receptor can be separated
from the intracellular domain by the hinge, CH2 and CH3 domains of human
IgG1 (see AAR89441). The extracellular portion of the chimeric receptor
contains a fragment of CD4 (amino acids 1-394 or 1-200 of the CD4
sequence, see AAR89450 and AAR89451) which specifically recognizes and
binds HIV-infected cells, but does not mediate HIV infection. The
extracellular domain of the receptor is separated from the cell membrane
by 48 or 72 angstroms, or by one or more proteinaceous alpha-helices.
The cells expressing the receptor are preferably T cells, B cells,
neutrophils, or dendritic cells. The therapeutic cells expressing the
chimeric receptor are administered to a mammal to treat HIV infection

```

```

SQ Sequence 462 AA;
Query Match 59.5%; Score 2032.5; DB 2; Length 462;
Best Local Similarity 98.0%; Pred. No. 7.8e-102;
Matches 396; Conservative 2; Mismatches 3; Indels 3; Gaps 1;

QY 1 MNRGVPRHLILVQLALPRAATQGNKVYLGKGDVETLCTASQKKSIOFHMNKNQIK 60
D 1 MNRGVPRHLILVQLALPRAATQGNKVYLGKGDVETLCTASQKKSIOFHMNKNQIK 60
QY 61 ILGNQGSFLTKGSPSKLNDRADSRSLMDQGNFPLIINKLIEDSDTYICEVEDQKEEVOL 120
D 61 ILGNQGSFLTKGSPSKLNDRADSRSLMDQGNFPLIINKLIEDSDTYICEVEDQKEEVOL 120
QY 121 LVFGLTANSTPHLLQGGSLTLTLSPGSSPSVQCSPRKNTQGGKTLISVSQLELDQSG 180
D 121 LVFGLTANSTPHLLQGGSLTLTLSPGSSPSVQCSPRKNTQGGKTLISVSQLELDQSG 180
QY 181 TWCTVLONOKKVEFKIDIVLAFQKASSIYKKEGQVEFSFPLAFTVEKLTGSGELMW 240
D 181 TWCTVLONOKKVEFKIDIVLAFQKASSIYKKEGQVEFSFPLAFTVEKLTGSGELMW 240
QY 241 QAERASSSSKSWITFDLKNKEVSVRVTQDPKLGKGLPLHLTLPQALPOYAGSGNLTLA 300
D 241 QAERASSSSKSWITFDLKNKEVSVRVTQDPKLGKGLPLHLTLPQALPOYAGSGNLTLA 300
QY 301 LEATGTLHQBVLVWVRATQLOKNTLCEVWGPTSPKLMSTLKENKEAKVSKREKPVWV 360
D 301 LEATGTLHQBVLVWVRATQLOKNTLCEVWGPTSPKLMSTLKENKEAKVSKREKPVWV 360
QY 361 LNPEAGMOCILSDSGOVLLESNIKVLPTWSTPVPCEAPBPKSC 404
D 361 LNPEAGMOCILSDSGOVLLESNIKVLPTWSTPVPCEAPBPKSC 404

RESULT 47
AAM02214
ID AAM02214 standard; protein; 462 AA.
AC AAM02214;
DT 11-NOV-1996 (first entry)
DE CD4:Fc receptor gamma chain chimaeric receptor.
KM Chimaeric receptor; cellular immunity; adoptive immunotherapy; CD4;
human immunodeficiency virus type 1; HIV-1; AIDS; therapy;
Fc receptor gamma chain; cytotoxic T lymphocyte; CTL.
OS Homo sapiens.
PN WO9625953-A1.
PD 29-AUG-1996.
PF 25-JAN-1996; 96WO-US001056.
PR 24-FEB-1995; 95US-00394176.

Location/Qualifiers
1..393 /label= "Extracellular domain"
394..397 /label= "linker"
/note= "encoding DNA contains a BamHI site used for
fusion construction"
398..462 /note= "region of fusion derived from gamma chain,
preferred signal-transducing portions for constructs of
the invention are amino acids 421-462 and 402-419"
400..462 /label= "Transmembrane+intracellular domains"

```

```

XX (GEHO ) GEN HOSPITAL CORP.
PA Seed B, Romeo C, Kolanus W;
XX WPI; 1996-402134/40.
XX N-PSDB; AAT36759.
XX
XX Direction of cellular immune response using therapeutic cell expressing 2
XX chimeric receptors - comprising region binding to target cell and region
XX that signals target cell destruction, or CD28 region, partic. for
XX eliminating HIV-infected cells.
XX
XX Claim 7; Page 76; 120pp; English.
XX
XX A chimeric receptor (AAW0214) comprises the extracellular domain of an
XX engineered form of the CD4 cellular receptor for HIV and the
XX transmembrane and intracellular regions, including the cytosolic signal-
XX transducing portion, of the human Fc receptor gamma chain; the region of
XX the fusion is shown in AAW0223. It can be obt'd. by inserting a gene
XX fusion (AAT36759) into a vaccinia virus vector and expressing in host
XX cells. Chimeric receptors comprising CD4 fused to Fc receptor gamma or T
XX -cell receptor zeta (see also AAW02213) or eta (AAW02215) chains are
XX capable of directing cytotoxic T lymphocytes to specifically recognise
XX and kill cells expressing HIV gp120, thus providing a therapy for AIDS
XX
XX Sequence 462 AA;
SQ
Query Match          59.5%; Score 2032.5; DB 2; Length 462;
Best Local Similarity 98.0%; Pred. No. 7.8e-102;
Matches 396; Conservative 2; Mismatches 3; Indels 3; Gaps 1;
QY 1 MNRGVFRRHLIVLQALPAATQGNKVVIGKKGDTVELTCTASQKKSIOFHWKNSNQIK 60
DB 1 MNRGVFRRHLIVLQALPAATQGNKVVIGKKGDTVELTCTASQKKSIOFHWKNSNQIK 60
QY 61 ILGNQGSFLTKGPSKLNDRADRSRSLMDQGNFPLIINKLIEDSDTYICEVEDQKEVOL 120
DB 61 ILGNQGSFLTKGPSKLNDRADRSRSLMDQGNFPLIINKLIEDSDTYICEVEDQKEVOL 120
QY 121 LVFGLTANSDTHLLOGOSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSDTHLLOGOSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
QY 181 TWICTVLQONQKVEFKIDIVLAFOKASSIVYKKEGQVEFSPPLAFTVEKLTGSGELMW 240
DB 181 TWICTVLQONQKVEFKIDIVLAFOKASSIVYKKEGQVEFSPPLAFTVEKLTGSGELMW 240
QY 241 QABRASSKSWITFDLKNKEVSVKRVTOPDKLQMGKKLPLHLTLPLPALPOYAGSGNLTLLA 300
DB 241 QABRASSKSWITFDLKNKEVSVKRVTOPDKLQMGKKLPLHLTLPLPALPOYAGSGNLTLLA 300
QY 301 LEAKTGKLEHENVLVVWRATOLQKNLTCCEWGPSPKMLSLKLENKEAKVSREKRPVWV 360
DB 301 LEAKTGKLEHENVLVVWRATOLQKNLTCCEWGPSPKMLSLKLENKEAKVSREKRPVWV 360
QY 361 LNPEAGMOCCLSDSGOVLLESNIKVLPTWSTPVPAPBPSC 404
DB 361 LNPEAGMOCCLSDSGOVLLESNIKVLPTWSTPVPVHA---DPOLC 401

```

```

KM CD4; B cell receptor; Fc receptor; pathogen; bacterial; fungal;
KM protozoan; viral.
XX Synthetic.
XX Mus sp.
XX US843728-A.
XX
XX 01-DEC-1998.
XX
XX 05-APR-1995; 95US-00417495.
XX
XX 07-MAR-1991; 91US-00665961.
XX 06-MAR-1992; 92US-00847566.
XX 28-FEB-1994; 94US-00203866.
XX
XX (GEHO ) GEN HOSPITAL CORP.
XX
XX Romeo C, Kolanus W, Seed B;
XX
XX WPI; 1999-044582/04.
XX N-PSDB; AAV70158.
XX
XX Membrane-bound chimeric receptors - comprising extracellular portion
XX which recognises and binds a target cell and an intracellular portion of
XX e.g. a T-cell receptor.
XX
XX Example 2; Col 43-46; 57pp; English.
XX
XX The present invention describes DNA encoding a membrane-bound chimeric
XX receptor comprising: (a) an extracellular portion that specifically
XX recognises and binds a target cell or a target infective agent; and (b)
XX an intracellular portion of a T-cell receptor CD3, zeta or eta
XX polypeptide, a B-cell receptor polypeptide or an Fc receptor polypeptide.
XX The present sequence represents a chimeric receptor containing the mouse
XX gamma polypeptide. Cells expressing chimeric receptors of the present
XX invention can be administered to mammals in order to destroy pathogens
XX (e.g. bacteria, fungi, protozoa or viruses, especially HIV), cancer cells
XX or autoimmune-generated cells
XX
XX Sequence 462 AA;
SQ
Query Match          59.5%; Score 2032.5; DB 2; Length 462;
Best Local Similarity 98.0%; Pred. No. 7.8e-102;
Matches 396; Conservative 2; Mismatches 3; Indels 3; Gaps 1;
QY 1 MNRGVFRRHLIVLQALPAATQGNKVVIGKKGDTVELTCTASQKKSIOFHWKNSNQIK 60
DB 1 MNRGVFRRHLIVLQALPAATQGNKVVIGKKGDTVELTCTASQKKSIOFHWKNSNQIK 60
QY 61 ILGNQGSFLTKGPSKLNDRADRSRSLMDQGNFPLIINKLIEDSDTYICEVEDQKEVOL 120
DB 61 ILGNQGSFLTKGPSKLNDRADRSRSLMDQGNFPLIINKLIEDSDTYICEVEDQKEVOL 120
QY 121 LVFGLTANSDTHLLOGOSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSDTHLLOGOSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
QY 181 TWICTVLQONQKVEFKIDIVLAFOKASSIVYKKEGQVEFSPPLAFTVEKLTGSGELMW 240
DB 181 TWICTVLQONQKVEFKIDIVLAFOKASSIVYKKEGQVEFSPPLAFTVEKLTGSGELMW 240
QY 241 QABRASSKSWITFDLKNKEVSVKRVTOPDKLQMGKKLPLHLTLPLPALPOYAGSGNLTLLA 300
DB 241 QABRASSKSWITFDLKNKEVSVKRVTOPDKLQMGKKLPLHLTLPLPALPOYAGSGNLTLLA 300
QY 301 LEAKTGKLEHENVLVVWRATOLQKNLTCCEWGPSPKMLSLKLENKEAKVSREKRPVWV 360
DB 301 LEAKTGKLEHENVLVVWRATOLQKNLTCCEWGPSPKMLSLKLENKEAKVSREKRPVWV 360
QY 361 LNPEAGMOCCLSDSGOVLLESNIKVLPTWSTPVPAPBPSC 404
DB 361 LNPEAGMOCCLSDSGOVLLESNIKVLPTWSTPVPVHA---DPOLC 401

```

ID	AA002215	standard; protein; 532 AA.
XX	AA002215;	
XX	16-OCT-2003	(revised)
XX	11-NOV-1996	(first entry)
XX	CD4:T-cell receptor eta chain chimaeric receptor.	
XX	Chimaeric receptor; cellular immunity; adoptive immunotherapy; CD4;	
XX	human immunodeficiency virus type 1; HIV-1; AIDS; therapy;	
XX	T-cell receptor eta chain; cytotoxic T lymphocyte; CTL.	
OS	Homo, sapiens.	
OS	Mus sp.	
OS	Chimeric.	
XX	Key	Location/Qualifiers
XX	Domain	1..393
XX		/label= "Extracellular domain"
XX		/note= "CD4 extracellular domain"
XX	Region	394..396
XX		/label= "Linker"
XX		/note= "encoding DNA contains a BamHI site used for fusion construction"
XX	Region	397..532
XX		/note= "region of fusion derived from eta chain, preferred signal-transducing portions for constructs of the invention are amino acids 421-532, 423-455, 438-455, 461-494, 494-528 or 400-420"
XX	Domain	400..437
XX		/label= "Transmembrane domain"
XX		/note= "eta chain transmembrane domain"
XX	Domain	438..575
XX		/label= "Intracellular domain"
XX		/note= "eta chain intracellular domain"
XX	MO3625953-A1.	
XX	29-AUG-1996.	
XX	25-JAN-1996;	96WO-US001056.
XX	24-FEB-1995;	95US-00394176.
XX	(GENO) GEN HOSPITAL CORP.	
XX	Seed B, Romeo C, Kolanus W;	
XX	WPI; 1996-402134/40.	
XX	N-PSDB; AAT36760.	
XX	Direction of cellular immune response using therapeutic cell expressing 2	
XX	chimaeric receptors - destricting region binding to target cell and region	
XX	et al; signals target cell destruction, or CD8 region, partic. for	
XX	eliminating HIV-infected cells.	
XX	Claim 7; Page 77-78; 120pp; English.	
XX	A chimeric receptor (AA002215) comprises the extracellular domain of an	
XX	engineered form of the CD4 cellular receptor for HIV and the	
XX	transmembrane and intracellular regions, including the cytoleptic signal-	
XX	transducing portion, of the mouse T-cell receptor eta chain. It can be	
XX	obt. by inserting a gene fusion (AAT36760) into a vaccinia virus vector	
XX	and expression in host cells. Chimaeric receptors comprising CD4 fused to	
XX	eta, eta (see also AA002213) or Fc receptor gamma (see also AA002214)	
XX	chains are capable of directing cytotoxic T lymphocytes to specifically	
XX	recognise and kill cells expressing HIV gp120, thus providing a therapy	
XX	for AIDS. (Updated on 16-Oct-2003 to standardise OS field)	

XX	Sequence	532 AA:
XX	Query Match	59.5%; Score 2032.5; DB 2; Length 532;
XX	Best Local Similarity	98.0%; Pred. No. 9,1e-102;
XX	Matches 396; Conservative	2; Mismatches 3; Indels 3; Gaps 1;
QY	1	MNRGVPPRHLLLVQLALLPAATGNNVVLGKKDVTVELTCTASQKKSIOFHMKNSNOIK 60
DB	1	MNRGVPPRHLLLVQLALLPAATGNNVVLGKKDVTVELTCTASQKKSIOFHMKNSNOIK 60
QY	61	ILGNQGSFLTYGPKSLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEEYQL 120
DB	61	ILGNQGSFLTYGPKSLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEEYQL 120
QY	121	LVFGLTANSDDTHLLQGGSLTTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDSG 180
DB	121	LVFGLTANSDDTHLLQGGSLTTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDSG 180
QY	181	TWCTCTVLOQNKKEVFKDIDIVLAQKASSIYYKKEGQVESPFLATVETLTSSGELMW 240
DB	181	TWCTCTVLOQNKKEVFKDIDIVLAQKASSIYYKKEGQVESPFLATVETLTSSGELMW 240
QY	241	QAECASSKSMITFDLANKKESVVRKYTDOPDLQMGKTLPLHLTLPOALPOYAGSGNLTIA 300
DB	241	QAECASSKSMITFDLANKKESVVRKYTDOPDLQMGKTLPLHLTLPOALPOYAGSGNLTIA 300
QY	301	LEATKGLKHOEVLVVMRATOLQKNLTCCEVWGPTSPKLMLSKLENKBAKVSKEKPYMV 360
DB	301	LEATKGLKHOEVLVVMRATOLQKNLTCCEVWGPTSPKLMLSKLENKBAKVSKEKPYMV 360
QY	361	LNPAAGMWQCLSDSGOYLLESNIKIVLPYMWSTPLPCAPAEPKXC 404
DB	361	LNPAAGMWQCLSDSGOYLLESNIKIVLPYMWSTPLPCAPAEPKXC 404
XX	RESULT 50	
XX	AAR89450	
XX	AAR89450	standard; peptide; 398 AA.
XX	AC	AAR89450;
XX	DT	26-SEP-1996 (first entry)
XX	DE	CD4 DI-D4 domains.
XX	KW	CD7; transmembrane domain; chimeric receptor; CD5; CD34; CH2; CH3; IgG1;
XX	KW	human; CD4; HIV; proteinnaceous alpha-helix; T cell; B cell; neutrophil;
XX	OS	dendritic cell; therapy; mammal; infection.
XX	PN	Homo sapiens.
XX	PD	WO9603883-A1.
XX	PF	26-JUL-1995; 95WO-US009468.
XX	PR	02-AUG-1994; 94US-00284391.
XX	PA	24-FEB-1995; 95US-00394388.
XX	PI	(GEHO) GEN HOSPITAL CORP.
XX	DR	Seed B, Banapour B, Romeo C, Kolanus W;
XX	DR	WPI; 1996-128034/13.
XX	PS	N-PSDB; AAT10797.
XX	PS	Membrane-bound chimeric receptor comprising extracellular portion
XX	PS	including CD4 fragment - cells expressing receptor can be used for
XX	PS	treatment of HIV infection.
XX	PS	Example 10; Fig 23; 134dp; English

XX This sequence represents the D1-D4 domains of CD4. This sequence is
CC included in the membrane bound proteinaceous chimeric receptor of the
CC invention. The extracellular portion of the chimeric receptor contains a
CC fragment of CD4 (amino acids 1-394 or 1-200 of the CD4 sequence) which
CC specifically recognizes and binds HIV-infected cells, but does not
CC mediate HIV infection. The extracellular domain of the receptor is
CC separated from the cell membrane by 48 or 72 angstroms, or by one or more
CC proteinaceous alpha-helices. The transmembrane region of the chimeric
CC receptor contains a portion of the CD7, CD5 or CD34 transmembrane domain.
CC Alternatively, the extracellular portion of the receptor can also be
CC separated from the intracellular domain by the hinge, CH2 and CH3 domains
CC of human IgG1. The cells expressing the receptor are preferably T cells,
CC B cells, neutrophils, or dendritic cells. The therapeutic cells
CC expressing the chimeric receptor are administered to a mammal to treat
CC HIV infection
XX
SQ Sequence 398 AA;

Query Match	59.4%	Score 2029;	DB 2;	Length 358;
Best Local Similarity	100.0%	Pred. No. 1e-101;		
Matches 554; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	I	INMRGVPFRLLLVLOLALLPAATQGNKVYLGGKGDVVELTCTASQOKKSIOFPHKNSNQIK	60
Db	1	NMRGVFRLLLVLOLALLPAATQGNKVYLGGKGDVVELTCTASQOKKSIOFPHKNSNQIK	60
QY	61	IILGNQGSFLTQKPSKLNDRADRSRSLMDQGNFLLIKNLKIEDSDPIYICEVEOKEEVL	120
Db	61	IILGNQGSFLTQKPSKLNDRADRSRSLMDQGNFLLIKNLKIEDSDPIYICEVEOKEEVL	120
QY	121	LVFGITANSDTHLLOQOSITLTLESPPGSSPSVQCRSPKGNKIQSGKITLSVSOLELQDSG	180
Db	121	LVFGITANSDTHLLOQOSITLTLESPPGSSPSVQCRSPKGNKIQSGKITLSVSOLELQDSG	180
QY	181	TMTCTVLQNOQKVEFKIDIVLVAFOQASSIVYKKEBQVEFSPPLAFYVEKLTGSGELMW	240
Db	181	TMTCTVLQNOQKVEFKIDIVLVAFOQASSIVYKKEBQVEFSPPLAFYVEKLTGSGELMW	240
QY	241	QAEKASSSKSMITFDLKNKEVSVKRTQDPKLOMGKKLPHLTLPOLPOLYAGSGNITTLA	300
Db	241	QAEKASSSKSMITFDLKNKEVSVKRTQDPKLOMGKKLPHLTLPOLPOLYAGSGNITTLA	300
QY	301	LEAKTGKTHQEVNLYVMRATOLQKNLTCEVWGPSTSEKMLSLKLENKEAVSRREKRVWV	360
Db	301	LEAKTGKTHQEVNLYVMRATOLQKNLTCEVWGPSTSEKMLSLKLENKEAVSRREKRVWV	360
QY	361	LNPEAGMOCCLSDSGVLLSENIKYLPTWSTPV	394
Db	361	LNPEAGMOCCLSDSGVLLSENIKYLPTWSTPV	394

RESULT 51
AAB07769
ID AAB07769 standard; protein; 458 AA.

AC AAB07769;

DT 07-NOV-2000 (first entry)

DE DNA encoding a human T4 glycoprotein.

Human; T4 glycoprotein; human immunodeficiency virus; HIV; KW

KW envelope glycoprotein; AIDS; virus binding.

Homo sapiens.

FH	Key	Location/Qualifiers
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

FT Peptide

```
FT      /note= "leader sequence"
EE      000 000
```

PT	Modified-site	296. .298
ET		

ET	Modified-air	325	327
ET		/note="N	

FT	/note= "N-linked glycosylation site"
FT	398. .420
FT	/note= "transmembrane domain"
FT	421. .458
FT	/note= "cytoplasmic domain"

PN US6093539-A.

PD 25-JUL-2000.

PF 06-JUN-1995; 95US-00466368.

PR 21-AUG-1986; 86US-00898587.

PR 06-JUL-1992; 92US-00909021.

XX

XX

XX

DR N-PSDB; AAA59352.

PT New isolated nucleic

PT for treating a subject

XX

XX
XX
XX

soluble polypeptide

CC envelope glycoprotein

CC a therapeutic agent,

CC treating human AIDS.

CC binding, or for dete

3 XX

111

Best Local Similarity

2.1. INTRODUCTION

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

Q: 21 TWOSEETK

11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527
 528
 529
 530
 531
 532
 533

101 TYPETANENT

131 TVECTANENT

191 TWENTY-ONE

101 TWISTED ONION

QY	241	QARASSSSKSWITFDLKNKEVS VKRVTDOPKLOMGKQJLHLTLPOLPOAGSGNITTLA	300
Db	241	QARASSSSKSWITFDLKNKEVS VKRVTDOPKLOMGKQJLHLTLPOLPOAGSGNITTLA	300
QY	301	LEAKGKGLHOEVLNVLVMPATOLQOKLTCGVWGPFSKMLSLKLENKEAVYSRREKEVVM	360
Db	301	LEAKGKGLHOEVLNVLVMPATOLQOKLTCGVWGPFSKMLSLKLENKEAVYSRREKEVVM	360

Qy 361 LNPEAGMWCCLSDSGVLLSNIKVLPTWSTPV 394
 Db 361 LNPEAGMWCCLSDSGVLLSNIKVLPTWSTPV 394

RESULT 52

AAR78673
 ID AAR78673 standard; protein; 398 AA.

XX AAR78673;

XX 12-APR-1996 (first entry)

DE CD4 domains D1-D4.

KM Chimeric receptor; CD4; T-cell receptor; HIV; cytolysis;
 KM human immunodeficiency virus; adoptive immunotherapy.

XX Homo sapiens.

XX MO9521528-A1.

XX 17-AUG-1995.

XX 12-JAN-1995; 95WO-US000454.

XX 14-FEB-1994; 94US-00195395.

XX 02-AUG-1994; 94US-00284391.

XX (GCHO) GEN HOSPITAL CORP.

XX Seed B, Banapour B, Romeo C, Kolanus W;

XX WPI; 1995-292893/38.

XX DR N-PSDB; AAQ96103.

XX Target cytolysis of HIV-infected cells - by chimeric CD4 receptor-bearing cells.

XX Example 10; Fig 23; 118pp; English.

XX Extracellular domains D1-D4 (AAR78673) or D1-D2 (AAR78674) of human CD4

XX are used in the construction of chimeric receptors utilised in the

XX targeted cytolysis of cells expressing HIV envelope proteins on their

XX surface. The chimeric receptors comprise the extracellular domain (pref.

XX amino acids 1-394 or 1-200) of CD4 linked to an intracellular portion,

XX e.g. of T-cell receptor protein zeta

XX Sequence 398 AA;

XX Query Match 59.3%; Score 2026; DB 2; Length 398;

XX Best Local Similarity 99.7%; Pred. No. 1.5e-101;

XX Matches 393; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MNRGVPFHLILVQLALLPATQGNKVVLLGKKGDTVELTGASOKSIQPHMKNNOIX 60

Db 1 MNRGVPFHLILVQLALLPATQGNKVVLLGKKGDTVELTGASOKSIQPHMKNNOIX 60

Qy 61 ILGNQGSFLTKGPSKLRADSRSLMDQGNFPLIKNLKTEDSDTYICEVEDQKEEVOL 120

Db 61 ILGNQGSFLTKGPSKLRADSRSLMDQGNFPLIKNLKTEDSDTYICEVEDQKEEVOL 120

Qy 121 LVFGILTANSDFHLLQGSLLTLTSSPPSSPSVQCRSRGNKIQGGKTLVSQLELDQSG 180

Db 121 LVFGILTANSDFHLLQGSLLTLTSSPPSSPSVQCRSRGNKIQGGKTLVSQLELDQSG 180

Qy 181 TMTCTVLONOKKVEFKIDIVLAFKASSIYKKKGGEQVRSFPLAFVETKLTGSGELMW 240

Db 181 TMTCTVLONOKKVEFKIDIVLAFKASSIYKKKGGEQVRSFPLAFVETKLTGSGELMW 240

Qy 301 LEAKTKLHQBVLNVRATOLQKLTCEVWGPTSPKLMSTLKENBAKYSKREKPVWV 360
 Db 301 LEAKTKLHQBVLNVRATOLQKLTCEVWGPTSPKLMSTLKENBAKYSKREKPVWV 360

RESULT 53

AAB19509
 ID AAB19509 standard; protein; 416 AA.

XX AAB19509;

XX 09-JAN-2001 (first entry)

DE CD4-IGM fusion protein CH4Mmu.

KM CD4; IGM; human; CD4Mmu; fusion protein; immunoglobulin; HIV; SIV; gp120;

XX therapy; diagnosis.

XX Homo sapiens.

XX Key Location/Qualifiers

XX Protein 1..395

XX FT /note= "CD4 extracellular region"

XX FT Protein 400..416

XX FT /note= "IGM heavy chain partial sequence"

XX US6117656-A.

XX 12-SEP-2000.

XX 07-JUN-1995; 95US-00479353.

XX 22-JAN-1988; 88US-00147351.

XX 23-JAN-1989; 89US-00299596.

XX 09-JUN-1992; 92US-00896781.

XX 12-APR-1993; 93US-00057952.

XX 04-FEB-1994; 94US-00191708.

XX (GCHO) GEN HOSPITAL CORP.

XX Seed B;

XX WPI; 2000-586558/55.

XX DR N-PSDB; AAA50662.

XX CD4-immunoglobulin fusion proteins, useful for targeting gp120 of HIV or

XX SIV.

XX Example 1; Col 41-50; 39pp; English.

XX The present sequence is that of fusion protein CD4Mmu comprising the

XX extracellular portion of CD4, which binds to HIV gp120, linked at its C-

XX terminus to the human IGM heavy chain. To obtain the fusion protein, DNA

XX encoding CD4 was linked to IGM DNA at the MscI site upstream of the Cbl

XX region (see AAA50662). Fusion protein CD4Mmu and a nucleic acid encoding

XX it are claimed. Also claimed are a vector comprising the nucleic acid,

XX and a method of producing the fusion protein in secreted form using a

XX transformed host cell. The fusion protein may further comprise a

XX therapeutic agent, radiolabel or NMR imaging agent. The fusion protein

XX can be administered to an animal (including humans) for treatment of HIV

XX or SIV infection, and can also be used in assays for HIV or SIV, imaging

XX and tissue stains. IGM fusion proteins such as CD4Mmu provide complement-

XX mediated immunity

XX Sequence 416 AA;

XX Query Match 59.3%; Score 2026; DB 3; Length 416;

XX Best Local Similarity 94.6%; Pred. No. 1.6e-101;

Matches 401; Conservative 0; Mismatches 7; Indels 16; Gaps 3;

QY 1 MNRGVPRRLILVLOALIPAAATQGNKVILGKKGDVVELTCTASQKKSIOFHMKNSNQIK 60
 DB 1 MNRGVPRRLILVLOALIPAAATQGNKVILGKKGDVVELTCTASQKKSIOFHMKNSNQIK 60
 QY 61 ILGNQSFLLTKGPKSKLNDRADSRSLMDQGNFPLIIKNKIEDSDTYICEVEDQKEEVL 120
 DB 61 ILGNQSFLLTKGPKSKLNDRADSRSLMDQGNFPLIIKNKIEDSDTYICEVEDQKEEVL 120
 QY 121 LVFGLTANSPTLHLOQSLTTLTLESPGSSPVQCRSPRGKNIQGGKTLVSQLELDG 180
 DB 121 LVFGLTANSPTLHLOQSLTTLTLESPGSSPVQCRSPRGKNIQGGKTLVSQLELDG 180
 QY 181 TWCTCTVLQNKQKVEFKIDIVLAFOKASSIVYKKEGQVEFSPLAFYVEKLTGSGELW 240
 DB 181 TWCTCTVLQNKQKVEFKIDIVLAFOKASSIVYKKEGQVEFSPLAFYVEKLTGSGELW 240
 QY 241 QAERASSSKSWITFDLKNKEVSVKRVTPDKLQMGKLLPLHLTLPALPOYAGSGNLTIA 300
 DB 241 QAERASSSKSWITFDLKNKEVSVKRVTPDKLQMGKLLPLHLTLPALPOYAGSGNLTIA 300
 QY 301 LEAKTGKLEHVEVNLVVMRATQLOKNLTCEVWGFTSPKMLSLKLENKEAVSKREKPVWV 360
 DB 301 LEAKTGKLEHVEVNLVVMRATQLOKNLTCEVWGFTSPKMLSLKLENKEAVSKREKPVWV 360
 QY 361 LNPEAGMOCCLSDSGOVLLESNIKVLPTWSTPVPCAPAPKSCDKHTTPELLGSPV 419
 DB 361 LNPEAGMOCCLSDSGOVLLESNIKVLPTWSTPV-----HADPE--GSASAP 405
 QY 420 FLFP 423
 DB 406 TLFP 409

RESULT 54
 AAR06374
 ID AAR06374 standard; protein; 400 AA.
 XX
 AC AAR06374;
 XX
 DT 31-OCT-2002 (revised)
 DT 20-DEC-1990 (first entry)
 XX
 DE Truncated form of soluble T4 encoded by pBG381.
 XX
 KM plasmid pBG381; soluble T4 protein; AIDS; ARC; HIV.
 XX
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..23
 FT /label= "secretory signal"
 FT /note= "hydrophobic"
 FT 24..117
 FT /label= "extracellular"
 FT /note= "homology to V-regions"
 FT 118..132
 FT /label= "extracellular"
 FT /note= "homology to J-regions"
 FT 133..397
 FT /label= "extracellular"
 FT /note= "glycosylated"
 XX
 XX WO9008198-A.
 XX
 XX PD 26-JUL-1990.
 XX
 XX PF 18-JAN-1989; 89US-00300096.
 XX
 XX PR 18-JAN-1989; 89US-00300096.
 XX
 PA (HARD) HARVARD COLLEGE.

XX
 PI Letvin NA;
 XX
 DR WPI; 1990-254040/33.
 DR N-PSDB; AAO05608.
 XX
 PT Treating or preventing AIDS, ARC or HIV infection - by administering an
 immunologically effective amt. of soluble T4 protein.
 PS Disclosure; Fig 2; 121pp; English.
 XX
 CC T4-encoding plasmid pBG381 was used to transform Chinese Hamster Ovary
 CC cells for the production of truncated T4. Soluble T4 is produced by
 CC virtue of the removal of the transmembrane and cytoplasmic domains. The
 CC soluble forms may be modified to increase their immunogenicity by
 CC addition of an adjuvant such as incomplete Freund's adjuvant. The T4
 CC interferes with HIV/T4 interaction and elicits anti-soluble T4 antibody
 CC production. See also AAO05607. (Updated on 31-OCT-2002 to add missing OS
 CC field.)
 XX
 SQ Sequence 400 AA;
 Query Match 59.2%; Score 2021; DB 2; Length 400;
 Best Local Similarity 99.7%; Pred. No. 2.8e-101;
 Matches 393; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MNRGVPRRLILVLOALIPAAATQGNKVILGKKGDVVELTCTASQKKSIOFHMKNSNQIK 60
 DB 1 MNRGVPRRLILVLOALIPAAATQGNKVILGKKGDVVELTCTASQKKSIOFHMKNSNQIK 60
 QY 61 ILGNQSFLLTKGPKSKLNDRADSRSLMDQGNFPLIIKNKIEDSDTYICEVEDQKEEVL 120
 DB 61 ILGNQSFLLTKGPKSKLNDRADSRSLMDQGNFPLIIKNKIEDSDTYICEVEDQKEEVL 120
 QY 121 LVFGLTANSPTLHLOQSLTTLTLESPGSSPVQCRSPRGKNIQGGKTLVSQLELDG 180
 DB 121 LVFGLTANSPTLHLOQSLTTLTLESPGSSPVQCRSPRGKNIQGGKTLVSQLELDG 180
 QY 181 TWCTCTVLQNKQKVEFKIDIVLAFOKASSIVYKKEGQVEFSPLAFYVEKLTGSGELW 240
 DB 181 TWCTCTVLQNKQKVEFKIDIVLAFOKASSIVYKKEGQVEFSPLAFYVEKLTGSGELW 240
 QY 241 QAERASSSKSWITFDLKNKEVSVKRVTPDKLQMGKLLPLHLTLPALPOYAGSGNLTIA 300
 DB 241 QAERASSSKSWITFDLKNKEVSVKRVTPDKLQMGKLLPLHLTLPALPOYAGSGNLTIA 300
 QY 301 LEAKTGKLEHVEVNLVVMRATQLOKNLTCEVWGFTSPKMLSLKLENKEAVSKREKPVWV 360
 DB 301 LEAKTGKLEHVEVNLVVMRATQLOKNLTCEVWGFTSPKMLSLKLENKEAVSKREKPVWV 360
 QY 361 LNPEAGMOCCLSDSGOVLLESNIKVLPTWSTPV 394
 DB 361 LNPEAGMOCCLSDSGOVLLESNIKVLPTWSTPV 394

RESULT 55
 AAP81990
 ID AAP81990 standard; protein; 458 AA.
 XX
 AC AAP81990;
 XX
 DT 25-MAR-2003 (revised)
 DT 15-OCT-1990 (first entry)
 XX
 DE Clone pT4B encoded HIV T4 glycoprotein.
 XX
 KM Human Immunodeficiency Virus T4 envelope glycoprotein; AIDS; ss.
 XX
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..23
 FT /label= leader peptide

FT Region 24. .117
 FT /label= variable-like region 1
 FT Region 118. .134
 FT /label= joining-like region 1
 FT Region 135. .189
 FT /label= variable-like region 2
 FT Region 190. .205
 FT /label= joining-like region 2
 FT Region 206. .287
 FT /label= variable-like region 3
 FT Region 288. .309
 FT /label= joining-like region 3
 FT Modified-site 297. .299
 FT /label= putative N-glycosylation site
 FT Region 310. .377
 FT /label= variable-like region 4
 FT Modified-site 325. .327
 FT /label= putative N-glycosylation site
 FT Region 378. .397
 FT /label= joining-like region 4
 FT Region 398. .420
 FT /label= trans-membrane region
 FT Region 421. .458
 FT /label= cytoplasmic region
 FT
 PN M08801304-A.
 XX 25-FEB-1988.
 PD 20-AUG-1987; 87WO-US002050.
 PF 21-AUG-1986; 86US-00898587.
 PR (UYCO-) COLUMBIA UNIV.
 PA (MADD/) MADDON P J.
 XX Ltctman DR, Maddon PJ, Chase L, Axel R, Weiss R, McDougall JS;
 PI WPI; 1988-064019/09.
 XX DR N-PSDB; AAN80512.
 DR Nucleic acid encoding T4 glyco:protein - used for treatment of AIDS and
 XX PT producing antibodies for use as vaccine for immunisation against AIDS.
 XX PS Disclosure: Page 7; 12pp; English.
 XX T4 protein encoded by part of 3kb insert from human T cell library
 CC (PT4B). (Updated on 25-MAR-2003 to correct PA field.)
 CC
 XX
 SO Sequence 458 AA;
 Query Match 59.2%; Score 2021; DB 1; Length 458;
 Best Local Similarity 99.7%; Pred. No. 3.2e-101;
 Matches 393; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 241 QAERASSSKSWITFDLKNKEVSVKRVTDPKLQWKKLPLHLTLPOALPOYAGSGNLTIA 300
 Oy 301 LEAKTGKHOEVNLVWRATQLOKNTFCVWGPFSPLMLSLKLENKAVSKREKPVW 360
 Db 301 LEAKTGKHOEVNLVWRATQLOKNTFCVWGPFSPLMLSLKLENKAVSKREKAVW 360
 Oy 361 LNPEAGMWQCLSDSGVLLSNIKVLPTWSTPV 394
 Db 361 LNPEAGMWQCLSDSGVLLSNIKVLPTWSTPV 394
 RESULT 56
 AAP91369
 ID AAP91369 standard; protein; 458 AA.
 XX
 AC AAP91369;
 XX
 DT 24-OCT-2003 (revised)
 DT 25-MAR-2003 (revised)
 DT 08-JAN-1990 (first entry)
 XX
 DE T4 protein.
 XX
 KM T4 protein; human immunodeficiency virus; AIDS; clone pT4B.
 XX
 OS Homo sapiens; (human).
 XX
 FH Key Location/Qualifiers
 FT Domain 1. .23
 FT Domain 24. .117
 FT Region 41
 FT Region 109
 FT Domain 118. .134
 FT Domain 135. .189
 FT Region 184
 FT Domain 190. .204
 FT Domain 205. .286
 FT Binding-site 287. .309
 FT Domain 296. .298
 FT Binding-site 310. .376
 FT Binding-site 325. .327
 FT Region 328
 FT Region 370
 FT Domain 377. .397
 FT Domain 398. .420
 FT Domain 421. .458
 XX
 PN EP330227-A.
 XX
 PD 30-AUG-1989.
 XX
 XX 24-FEB-1989; 89EP-00103297.
 PF
 XX 24-FEB-1988; 88US-00160348.
 PR
 XX (UYCO-) COLUMBIA UNIV.
 PA (SMIK) SMITHKLINE BECKMAN CORP.
 PA (UYCO) UNIV COLUMBIA NEW YORK.
 PA (UYCO) UNIV COLUMBIA NEW YORK.
 XX
 PI Maddon PJ, Axel R, Sweet RW, Arthos J;
 XX
 DR WPI; 1989-250337/35.
 DR
 XX Soluble T4 polypeptide derivs. - inhibitors of extracellular and cell to
 PT cell spread of HIV used in prevention and treatment of AIDS.
 XX
 XX Claim 1; Fig 6; 73pp; English.
 PS
 XX T4 protein (AAN90619) inhibits extracellular and cell-to-cell spread of
 CC HIV. The therapeutic agent consists of amino acids +3--185 fused to +351-
 CC +369; +3--106 fused to +351--369; or +3--185. Also used to identify

CC inhibitors of T4+ interactions, as target carrier proteins, and to
 CC generate monoclonal antibodies. Above features are: Domain 1 (D); starting
 CC at the N-terminal) = leader; D2 = variable-like-1; D3 = joining-like-1;
 CC D4 = V2; D5 = J2; D6 = J3; D7 = V4; D8 = J4; D9 = transmembrane;
 CC D10 = cytoplasmic; Regions are extracellular cysteines; and the 2 sites
 CC are potential N-linked glycosylation sites. (Updated on 25-MAR-2003 to
 CC correct PA field.) (Updated on 24-OCT-2003 to standardise OS field)

XX Sequence 458 AA;

Query Match 59.2%; Score 2021; DB 1; Length 458;
 Best Local Similarity 99.7%; Pred. No. 3.2e-101;
 Matches 393; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MNRGVFPRHLLVLQALLPAAATQGNKVLGKKGDVETLTCTASQKKSIOFHKNQK 60
 DB 1 MNRGVFPRHLLVLQALLPAAATQGNKVLGKKGDVETLTCTASQKKSIOFHKNQK 60
 QY 61 ILGNQGSFLTKGSPSKLNDRADSRSLMDQGNFPLIIKNLKIEDSDTYICEVEDQKEEYOL 120
 DB 61 ILGNQGSFLTKGSPSKLNDRADSRSLMDQGNFPLIIKNLKIEDSDTYICEVEDQKEEYOL 120
 QY 121 LVFGLTANSDTHLLOQGSITLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
 DB 121 LVFGLTANSDTHLLOQGSITLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
 QY 181 TWCTCTVLOQOKKVEFKIDIVVLAFOKASSIVYKKEGEVFEFSPPLAFTVEKLTGSGELMW 240
 DB 181 TWCTCTVLOQOKKVEFKIDIVVLAFOKASSIVYKKEGEVFEFSPPLAFTVEKLTGSGELMW 240
 QY 241 QABRASSSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPLHLTLPALPOYAGSGNLTLLA 300
 DB 241 QABRASSSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPLHLTLPALPOYAGSGNLTLLA 300
 QY 301 LEAKTGKLEHVEVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKENKEAKVSKREKPVVW 360
 DB 301 LEAKTGKLEHVEVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKENKEAKVSKREKPVVW 360
 QY 361 LNPBAGMMQCLSDSGQVLLSNIKVLPTWSTPV 394
 DB 361 LNPBAGMMQCLSDSGQVLLSNIKVLPTWSTPV 394

RESULT 57

AA139826
 ID AAY39826 standard; protein; 458 AA.

XX AAY39826;
 AC
 XX
 DT 03-DEC-1999 (first entry)
 XX
 DE Soluble human T4 protein.
 XX
 KM Soluble T4 protein; sT4; human; HIV; binding inhibitor; T4+ cell; AIDS;
 KM vaccine; immunisation; therapy.
 OS Homo sapiens.
 XX
 OS
 XX
 PN US5958678-A.
 XX
 PD 28-SEP-1999.
 XX
 PF 12-DEC-1994; 94US-00354452.
 XX
 PR 21-AUG-1986; 86US-00898587.
 PR 11-JUN-1991; 91US-00713564.
 PR 06-JUL-1992; 92US-00909021.
 XX
 PA (UYCO) UNIV COLUMBIA NEW YORK.
 XX
 PI McDougal JS, Weiss R, Axel R, Littman DR, Maddon PJ, Chess L;
 XX WPL; 1999-561025/47.
 DR

DR N-PSDB; AA220695.

XX Human T4 protein inhibits HIV binding to T4 cells, useful for treating
 PT AIDS.
 PT

XX Example 3; Fig 6; 58pp; English.

XX This sequence represents the soluble human T4 protein of the invention.
 CC The soluble human T4 protein blocks the binding of HIV to T4+ cells and
 CC is therefore useful for the treatment of AIDS. Monoclonal antibodies
 CC against the T4 protein may be used as vaccines for immunising subjects
 CC against AIDS

XX Sequence 458 AA;

Query Match 59.2%; Score 2021; DB 2; Length 458;
 Best Local Similarity 99.7%; Pred. No. 3.2e-101;
 Matches 393; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MNRGVFPRHLLVLQALLPAAATQGNKVLGKKGDVETLTCTASQKKSIOFHKNQK 60
 DB 1 MNRGVFPRHLLVLQALLPAAATQGNKVLGKKGDVETLTCTASQKKSIOFHKNQK 60
 QY 61 ILGNQGSFLTKGSPSKLNDRADSRSLMDQGNFPLIIKNLKIEDSDTYICEVEDQKEEYOL 120
 DB 61 ILGNQGSFLTKGSPSKLNDRADSRSLMDQGNFPLIIKNLKIEDSDTYICEVEDQKEEYOL 120
 QY 121 LVFGLTANSDTHLLOQGSITLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
 DB 121 LVFGLTANSDTHLLOQGSITLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
 QY 181 TWCTCTVLOQOKKVEFKIDIVVLAFOKASSIVYKKEGEVFEFSPPLAFTVEKLTGSGELMW 240
 DB 181 TWCTCTVLOQOKKVEFKIDIVVLAFOKASSIVYKKEGEVFEFSPPLAFTVEKLTGSGELMW 240
 QY 241 QABRASSSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPLHLTLPALPOYAGSGNLTLLA 300
 DB 241 QABRASSSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPLHLTLPALPOYAGSGNLTLLA 300
 QY 301 LEAKTGKLEHVEVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKENKEAKVSKREKPVVW 360
 DB 301 LEAKTGKLEHVEVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKENKEAKVSKREKPVVW 360
 QY 361 LNPBAGMMQCLSDSGQVLLSNIKVLPTWSTPV 394
 DB 361 LNPBAGMMQCLSDSGQVLLSNIKVLPTWSTPV 394

RESULT 58

AA04032
 ID AAR04032 standard; protein; 2037 AA.

XX AAR04032;
 AC
 XX
 DT 25-MAR-2003 (revised)
 DT 31-OCT-2002 (revised)
 DT 29-MAY-1990 (first entry)
 XX
 DE Full length T4 encoded by plasmid pBG381.
 XX
 KM Soluble T4; pBG381; anti-retroviral agent; AIDS; ARC; HIV; AZT.
 XX
 OS Synthetic.
 XX
 PN W08911860-A.
 XX
 PD 14-DEC-1989.
 XX
 PF 08-JUN-1989; 89WO-US002453.
 XX
 PR 10-JUN-1988; 88US-00204645.
 PR 20-APR-1989; 89US-00341080.
 XX

PA (BIOJ) BIOGEN NV INC.
 PA (GHO) GEN HOSPITAL CORP.
 PA (BIOJ) BIOGEN INC.
 PA (BIOJ) BIOGEN INC.
 PI Fisher RA, Schooley RT, Hirsch MS, Johnson VA, Walker BD;
 DR WPI; 1990-007302/01.
 DR N-PSDB; AAQ03006.
 XX
 PT Combinations of soluble T4 protein and anti-retroviral agent - having
 PT synergistic activity in treatment and prevention of AIDS, arc and HIV
 PT infection.
 PS
 PS Disclosure; Fig 2; 100pp; English.
 XX
 CC X = stop codon. The sequence was deduced from the cDNA insert of pBG383.
 CC Soluble T4 constructs may be produced by truncating this sequence to give
 CC fragments from position 400 to 799, removing the transmembrane and
 CC intracytoplasmic domains whilst retaining the extracellular region
 CC responsible for HIV binding. The sol. T4 is combined with an anti-viral
 CC agent such as AZT. See also AAQ03005. (Updated on 31-OCT-2002 to add
 CC missing OS field.) (Updated on 25-MAR-2003 to correct PA field.)
 XX
 SQ Sequence 2037 AA;
 Query Match 59.2%; Score 2021; DB 2; Length 2037;
 Best Local Similarity 99.7%; Pred. No. 1.5e-100;
 Matches 393; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MNRGVPRHLLLVQLALPAAIQGNKVVLGKGGDTVELCTASQKSIQPHMKNNOIK 60
 DB 403 MNRGVPRHLLLVQLALPAAIQGNKVVLGKGGDTVELCTASQKSIQPHMKNNOIK 462
 QY 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNFPIIKNLKIEDSDTYICEVEDQKEEVOL 120
 DB 463 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNFPIIKNLKIEDSDTYICEVEDQKEEVOL 522
 QY 121 LVFGLTANSPTDHLQOGSLTLTLSPSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
 DB 523 LVFGLTANSPTDHLQOGSLTLTLSPSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 582
 QY 181 TWCTTVLQONKVEFKIDIVLAFQKASSIVYKKEGEQVEFSPLAFTVEKLTGSGELMW 240
 DB 583 TWCTTVLQONKVEFKIDIVLAFQKASSIVYKKEGEQVEFSPLAFTVEKLTGSGELMW 642
 QY 241 QAEBASSSSKSWITFDLKNKEVSVKRVTPDPLQMGKKLPLHLTLPOALPOYAGSGNLTIA 300
 DB 643 QAEBASSSSKSWITFDLKNKEVSVKRVTPDPLQMGKKLPLHLTLPOALPOYAGSGNLTIA 702
 QY 301 LEAKTGKHOEVNLVWMRATOLQNLTCFVWGPTSPKMLSLKLENKAKVSKREKAVW 360
 DB 703 LEAKTGKHOEVNLVWMRATOLQNLTCFVWGPTSPKMLSLKLENKAKVSKREKAVW 762
 QY 361 LNPEAGMOCCLSDSGOVLLESNIKVLPTWSTPV 394
 DB 763 LNPEAGMOCCLSDSGOVLLESNIKVLPTWSTPV 796
 RESULT 59
 AAR07641 ID AAR07641 standard; protein; 2050 AA.
 XX
 AC AAR07641;
 XX
 DT 31-OCT-2002 (revised)
 DT 20-DEC-1990 (first entry)
 XX
 DE Deduced sequence of pBG381 comprising truncated T4 glycoprotein.
 XX plasmid pBG381; soluble T4 protein; AIDS; ARC; HIV.
 XX
 OS Synthetic.

XX Key Location/Qualifiers
 FT 403, 803
 FT Protein /label= truncated soluble T4 glycoprotein
 XX
 XX W09008198-A.
 XX
 XX 26-JUL-1990.
 XX
 XX 18-JAN-1989; 89US-00300096.
 XX
 XX 18-JAN-1989; 89US-00300096.
 XX
 XX 18-JAN-1989; 89US-00300096.
 XX
 XX (HARD) HARVARD COLLEGE.
 XX
 PI Letvin NA;
 XX
 DR WPI; 1990-254040/33.
 DR N-PSDB; AAQ05608.
 XX
 PT Treating or preventing AIDS, ARC or HIV infection - by administering an
 PT immunologically effective amt. of soluble T4 protein.
 PS
 PS Disclosure; Fig 2; 121pp; English.
 XX
 CC Entire sequence translation of plasmid pBG381 used to transform Chinese
 CC Hamster Ovary cells for the production of soluble truncated T4.
 CC Transmembrane and cytoplasmic domain-encoding regions are deleted from
 CC the T4 COS to increase their immunogenicity by addition of an adjuvant such
 CC as incomplete Freund's adjuvant. The T4 interferes with HIV/T4
 CC interaction and elicits anti-soluble T4 antibody prodn. See also
 CC AAQ05607. (Updated on 31-OCT-2002 to add missing OS field.)
 XX
 SQ Sequence 2050 AA;
 Query Match 59.2%; Score 2021; DB 2; Length 2050;
 Best Local Similarity 99.7%; Pred. No. 1.5e-100;
 Matches 393; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MNRGVPRHLLLVQLALPAAIQGNKVVLGKGGDTVELCTASQKSIQPHMKNNOIK 60
 DB 403 MNRGVPRHLLLVQLALPAAIQGNKVVLGKGGDTVELCTASQKSIQPHMKNNOIK 462
 QY 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNFPIIKNLKIEDSDTYICEVEDQKEEVOL 120
 DB 463 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNFPIIKNLKIEDSDTYICEVEDQKEEVOL 522
 QY 121 LVFGLTANSPTDHLQOGSLTLTLSPSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
 DB 523 LVFGLTANSPTDHLQOGSLTLTLSPSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 582
 QY 181 TWCTTVLQONKVEFKIDIVLAFQKASSIVYKKEGEQVEFSPLAFTVEKLTGSGELMW 240
 DB 583 TWCTTVLQONKVEFKIDIVLAFQKASSIVYKKEGEQVEFSPLAFTVEKLTGSGELMW 642
 QY 241 QAEBASSSSKSWITFDLKNKEVSVKRVTPDPLQMGKKLPLHLTLPOALPOYAGSGNLTIA 300
 DB 643 QAEBASSSSKSWITFDLKNKEVSVKRVTPDPLQMGKKLPLHLTLPOALPOYAGSGNLTIA 702
 QY 301 LEAKTGKHOEVNLVWMRATOLQNLTCFVWGPTSPKMLSLKLENKAKVSKREKAVW 360
 DB 703 LEAKTGKHOEVNLVWMRATOLQNLTCFVWGPTSPKMLSLKLENKAKVSKREKAVW 762
 QY 361 LNPEAGMOCCLSDSGOVLLESNIKVLPTWSTPV 394
 DB 763 LNPEAGMOCCLSDSGOVLLESNIKVLPTWSTPV 796
 RESULT 60
 AAP93506 ID AAP93506 standard; protein; 394 AA.
 XX

```

AC AAP93506;
XX
XX 25-MAR-2003 (revised)
DT 02-JUN-1990 (first entry)
XX
XX Derived sequence of soluble T4 lymphocyte surface protein (sT4).
XX
XX Soluble T4 lymphocyte surface protein; sT4; AIDS therapy; AIDS diagnosis.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH Protein 26..394
FT Misc-difference 26..26
FT /note= "When sequence was determined by amino acid
FT sequencing, this residue was 'lys.'"
FT
FT Region
FT /note= "These residues are identical to those determined
FT by amino acid sequencing"
XX
XX EP313377-A.
XX
XX 26-APR-1989.
XX
XX 21-OCT-1988; 88EP-00309907.
XX
XX 23-OCT-1987; 87US-00112800.
XX
XX (SMIK ) SMITHKLINE BECKMAN CORP.
XX (SMIK ) SMITHKLINE BECKMAN CORP.
XX
XX Deen KC, Folenawaas GM, Inacker RH, Sweet RW;
XX WPI; 1989-124209/17.
XX DR N-PSDB; AAN90763.
XX
XX Purifying soluble recombinant T4 lymphocyte surface protein - from cell
XX culture by adsorption on cation exchanger, elution and treatment with
XX anion exchanger.
XX
XX Disclosure; Fig 1; 13pp; English.
XX
XX The coding sequence is derived from the published sequence of sT4. sT4 is
XX useful in the prevention and treatment of AIDS by inhibiting spread of
XX the virus. It can also be used as an inhibitor of T4+ cell function, as a
XX reagent for identifying inhibitors of T4+ cell interaction and to produce
XX diagnostic monoclonal antibodies. (Updated on 25-MAR-2003 to correct PA
XX field.) (Updated on 25-MAR-2003 to correct PI field.)
XX
XX Sequence 394 AA;
XX
Query Match 59.1%; Score 2018; DB 1; Length 394;
Best Local Similarity 99.5%; Pred. No. 4e-101;
Matches 392; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 MNRGVFRRHLVLVQLALLPAATQGNKVVIGKKGDTVELTCTASQKKSIOFHKNNSQIK 60
DB 1 MNRGVFRRHLVLVQLALLPAATQGNKVVIGKKGDTVELTCTASQKKSIOFHKNNSQIK 60
QY ILGNQGSFLLTKGSPKLNDRADSRSLMDQGNFLLIKNLKTIKESDVTICEVEQKEEVOL 120
DB 61 ILGNQGSFLLTKGSPKLNDRADSRSLMDQGNFLLIKNLKTIKESDVTICEVEQKEEVOL 120
QY 121 LVFGLTANSDTHLLQGSGLTLTLESPPGSSPSVQCRSPKNIQGGKTLISVQLQLDSCG 180
DB 121 LVFGLTANSDTHLLQGSGLTLTLESPPGSSPSVQCRSPKNIQGGKTLISVQLQLDSCG 180
QY 121 LVFGLTANSDTHLLQGSGLTLTLESPPGSSPSVQCRSPKNIQGGKTLISVQLQLDSCG 180
DB 121 LVFGLTANSDTHLLQGSGLTLTLESPPGSSPSVQCRSPKNIQGGKTLISVQLQLDSCG 180
QY 181 TWCTCTVQNGKQKVEFKIDIVLAFOKASSIVYKKEGQVEFSPPLATTVKLTGSGELMW 240
DB 181 TWCTCTVQNGKQKVEFKIDIVLAFOKASSIVYKKEGQVEFSPPLATTVKLTGSGELMW 240
QY 241 QAERASSSKSWITPDLKNKEVSVKRVQDPKLQWKGKLPHTLTPQALPYAGSGNLTLA 300

```

```

DB 241 QAERASSSKSWITPDLKNKEVSVKRVQDPKLQWKGKLPHTLTPQALPYAGSGNLTLA 300
QY 301 LEAKTGKLEHVEVNLVYMRATQLOKNIITCEVWGPSTSPKMLSLKLEKKAQVSKREPVVY 360
DB 301 LEAKTGKLEHVEVNLVYMRATQLOKNIITCEVWGPSTSPKMLSLKLEKKAQVSKREPVVY 360
QY 361 LNPEAGMWOCLLSDSGOVLLESNIKVLPTWSTPV 394
DB 361 LNPEAGMWOCLLSDSGOVLLESNIKVLPTWSTPV 394

RESULT 61
AAP91922
ID AAP91922 standard; protein; 402 AA.
XX
XX AAP91922;
XX
XX 25-MAR-2003 (revised)
XX 31-OCT-2002 (revised)
DT 14-MAY-1990 (first entry)
XX
XX Sequence of a secreted form of the CD4 adhesion (CD4T) polypeptide.
XX
XX CD4 variants; CD4T; gp120; plasmid pRKCDA; HIV-1; HTLV-IIIB.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH Misc-difference 25..26
FT /note= "signal processing site"
FT
FT Misc-difference 366
FT /note= "other forms of CD4T terminate here"
FT
FT Misc-difference 368
FT /note= "other forms of CD4T terminate here"
XX
XX EP314317-A.
XX
XX 03-MAY-1989.
XX
XX 03-OCT-1988; 88EP-00309194.
XX
XX 02-OCT-1987; 87US-00104329.
XX 28-SEP-1986; 88US-00250785.
XX
XX (GETH ) GENENTECH INC.
XX
XX Capon DJ, Gregory TJ;
XX WPI; 1989-131855/18.
XX DR N-PSDB; AAN90777.
XX
XX Compens. contg. adhesion variants - useful in therapy and diagnostics,
XX e.g. CD4 variants which are therapeutically useful for treating human
XX immuno-deficiency virus.
XX
XX Disclosure; Fig 1a-1c; 36pp; English.
XX
XX It may be capable of binding gp120. It may be fused with an
XX immunoglobulin constant domain, human transferrin, apolipoprotein,
XX albumin, ricin A chain or diphtheria toxin A. It may be used for
XX antiviral of immunomodulatory therapy particularly in treatment of HIV
XX infection. It may have variants by insertion, substitution of deletion in
XX non-functional regions. (Updated on 31-OCT-2002 to add missing OS field.)
XX (Updated on 25-MAR-2003 to correct PR field.) (Updated on 25-MAR-2003 to
XX correct PI field.)
XX
XX Sequence 402 AA;
XX
Query Match 59.1%; Score 2017; DB 1; Length 402;
Best Local Similarity 99.7%; Pred. No. 4.7e-101;
Matches 392; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MNRGVFRRHLVLVQLALLPAATQGNKVVIGKKGDTVELTCTASQKKSIOFHKNNSQIK 60

```

```

Db      1  MNRGVPRHLLLVYQLALLPAATQGNKVLGKKGDVLELCTASQKKSIOFHMKNQIK 60
Qy      61  ILNGQSFLLTKGPKLNDRADSRSLMDQGNFPLIINKLIEDSDTYICEVEDQKEVQL 120
Db      61  ILNGQSFLLTKGPKLNDRADSRSLMDQGNFPLIINKLIEDSDTYICEVEDQKEVQL 120
Qy      121  LVFGLTANSDTHLLOQGSLLTLTLESPGSSPVOCRSRPGKNIQSGKTLVSQLELDSG 180
Db      121  LVFGLTANSDTHLLOQGSLLTLTLESPGSSPVOCRSRPGKNIQSGKTLVSQLELDSG 180
Qy      181  TWTCVLOQNKKEVFKIDIVLAFQKASSIVYKKEGEVFPSPPLAFTVEKLTGSGELMW 240
Db      181  TWTCVLOQNKKEVFKIDIVLAFQKASSIVYKKEGEVFPSPPLAFTVEKLTGSGELMW 240
Qy      241  QAERASSSKSWITFDLKNKEVSVKRVYODPKLQMGKKLPHLTLPOALPOYAGSGNLTIA 300
Db      241  QAERASSSKSWITFDLKNKEVSVKRVYODPKLQMGKKLPHLTLPOALPOYAGSGNLTIA 300
Qy      301  LEAKTGKLEHENVLVVWRATOLQKNTLCEVWGPTSPKMLSLKLENKEAKVSKREKPYWV 360
Db      301  LEAKTGKLEHENVLVVWRATOLQKNTLCEVWGPTSPKMLSLKLENKEAKVSKREKPYWV 360
Qy      361  LNPEAGMWQCLLSDSGQVLLSNIKVLPTWSTP 393
Db      361  LNPEAGMWQCLLSDSGQVLLSNIKVLPTWSTP 393

RESULT 62
AAP94757
ID  AAP94757 standard; protein; 402 AA.
XX
XX  AAP94757;
XX
XX  25-MAR-2003 (revised)
XX  03-OCT-2002 (revised)
XX  28-JAN-1991 (first entry)
XX
DE  Sequence of a secreted form of the CD4 adhesion.
XX
XX  HIV; antiviral; therapy; diagnosis.
XX
XX  Homo sapiens.
XX
XX  Key 1. .25 location/Qualifiers
XX  Peptide /note="signal"
XX  Protein 26. .402
XX
XX  W08902922-A.
XX
XX  06-APR-1989.
XX
XX  03-OCT-1988; 88WO-US003414.
XX
XX  02-OCT-1987; 87US-00104329.
XX  28-SEP-1986; 88US-00250785.
XX
XX  (GETH ) GENENTECH INC.
XX
XX  Capon DJ, Gregory TJ;
XX
XX  WPI; 1989-114397/15.
XX  N-PSDB; AAN90734.
XX
XX  New nucleic acid sequences encoding adhesion, esp. CD 4, variants -
XX  partic. with trans-membrane domain inactivated or fused to other peptide,
XX  useful esp. for treating HIV infections.
XX
XX  Disclosure; Fig 1a-1c; 78pp; English.
XX
XX  The patent claims a nucleic acid encoding an aa sequence variant of an
XX  adhesion, which is pref. a CD4 polypeptide variant modified such that its

```

```

CC  transmembrane domain has been inactivated, either deleted or replaced by
CC  a sequence of hydrophilic hydropathy profile. The aa sequence variant of
CC  an adhesion may also be a fusion of CD4 with a 2nd polypeptide esp. one
CC  contg. a non-CD4 epitope; a signal sequence; a cgd. able to elicit a
CC  humoral immune response (viral polypeptide or allergen); or a human
CC  plasma protein of long plasma half-life. CD4 fusion proteins can have
CC  antiviral and immunomodulatory activity and are esp. useful for treating
CC  HIV infections regardless of genetic variation within the virus. They and
CC  antibodies raised against them can also be used diagnostically for
CC  assaying adhesions and their ligands. (updated on 03-OCT-2002 to add
CC  missing OS field.) (updated on 25-MAR-2003 to correct PR field.) (updated
CC  on 25-MAR-2003 to correct PA field.)
XX
XX  Sequence 402 AA;
XX
Query Match 59.1%; Score 2017; DB 1; Length 402;
Best Local Similarity 99.7%; Pred. No. 4,7e-101;
Matches 392; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1  MNRGVPRHLLLVYQLALLPAATQGNKVLGKKGDVLELCTASQKKSIOFHMKNQIK 60
Db      1  MNRGVPRHLLLVYQLALLPAATQGNKVLGKKGDVLELCTASQKKSIOFHMKNQIK 60
Qy      61  ILNGQSFLLTKGPKLNDRADSRSLMDQGNFPLIINKLIEDSDTYICEVEDQKEVQL 120
Db      61  ILNGQSFLLTKGPKLNDRADSRSLMDQGNFPLIINKLIEDSDTYICEVEDQKEVQL 120
Qy      121  LVFGLTANSDTHLLOQGSLLTLTLESPGSSPVOCRSRPGKNIQSGKTLVSQLELDSG 180
Db      121  LVFGLTANSDTHLLOQGSLLTLTLESPGSSPVOCRSRPGKNIQSGKTLVSQLELDSG 180
Qy      181  TWTCVLOQNKKEVFKIDIVLAFQKASSIVYKKEGEVFPSPPLAFTVEKLTGSGELMW 240
Db      181  TWTCVLOQNKKEVFKIDIVLAFQKASSIVYKKEGEVFPSPPLAFTVEKLTGSGELMW 240
Qy      241  QAERASSSKSWITFDLKNKEVSVKRVYODPKLQMGKKLPHLTLPOALPOYAGSGNLTIA 300
Db      241  QAERASSSKSWITFDLKNKEVSVKRVYODPKLQMGKKLPHLTLPOALPOYAGSGNLTIA 300
Qy      301  LEAKTGKLEHENVLVVWRATOLQKNTLCEVWGPTSPKMLSLKLENKEAKVSKREKPYWV 360
Db      301  LEAKTGKLEHENVLVVWRATOLQKNTLCEVWGPTSPKMLSLKLENKEAKVSKREKPYWV 360
Qy      361  LNPEAGMWQCLLSDSGQVLLSNIKVLPTWSTP 393
Db      361  LNPEAGMWQCLLSDSGQVLLSNIKVLPTWSTP 393

RESULT 63
AAV88328
ID  AAV88328 standard; protein; 394 AA.
XX
XX  AAV88328;
XX
XX  14-JUL-2000 (first entry)
XX
XX  T4 glycoprotein amino acid sequence.
XX
XX  sT4; glycoprotein; human immunodeficiency virus; HIV; block binding;
XX  AIDS; treatment; inhibit; cell to cell spread; infection; fusion.
XX
XX  Mammalia.
XX
XX  US5126433-A.
XX
XX  30-JUN-1992.
XX
XX  23-OCT-1987; 87US-00114244.
XX
XX  21-AUG-1986; 86US-00898587.
XX
XX  (UYCO ) UNIV COLUMBIA NEW YORK.
XX

```

PI Madden PJ, Chess L, Axel R, Weiss R, Littman DR, McDougal JS;
 XX MPI, 2000-348913/30.
 DR N-PSDB; AAA10906.
 XX
 XX Soluble T4 glycoprotein useful for prevention and treatment of acquired
 PT immunodeficiency syndrome and for screening inhibitors of human
 PT immunodeficiency viral binding.
 XX
 XX Disclosure; Col 11-16; 64pp; English.
 XX
 CC This sequence represents the full length amino acid sequence of
 CC glycosylated sT4 glycoprotein. Human immunodeficiency virus (HIV) uses
 CC sT4 as a target receptor on T cells. The invention relates to
 CC glycosylated sT4 which functions by blocking the binding of HIV to T4
 CC target cells, and can be used for the prophylaxis and treatment of AIDS
 CC patients. Administration of sT4 effectively inhibits the cell to cell
 CC spreading of HIV infection and also the fusion of HIV-infected T4 cells
 CC and non-infected T4 cells. The administration of T4 alleviates several
 CC symptoms associated with AIDS, and prevents the occurrence of new
 CC pathological changes. The sT4 glycoprotein is useful for the prophylaxis
 CC and treatment of patients with AIDS. It is also useful as a reagent to
 CC identify natural, synthetic or recombinant molecules which act as
 CC therapeutic agents or inhibitors of T4+ cell interactions and in
 CC diagnostic assays for detection T4 proteins or molecules
 CC
 XX
 SQ Sequence 394 AA;
 Query Match 59.0%; Score 2015; DB 3; Length 394;
 Best Local Similarity 99.5%; Pred. No. 5.8e-101;
 Matches 392; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MNRGVPFRHLILVQLALIPATQGNKVVGLKKGDVVELTCTASOKKSIQFHKNSNQIK 60
 DB 1 MNRGVPFRHLILVQLALIPATQGNKVVGLKKGDVVELTCTASOKKSIQFHKNSNQIK 60
 QY 61 ILGNQGSFLTKGPSKLNDRADRSRLMDQGNFPLIINKLKIEDSDTYICEVEDQKEEYVL 120
 DB 61 ILGNQGSFLTKGPSKLNDRADRSRLMDQGNFPLIINKLKIEDSDTYICEVEDQKEEYVL 120
 QY 121 LVFGLTANSDTHLLOQOSLTLTLESPPGSSPSVQCHSPRGKNIQGGKTLVSQLELDQSG 180
 DB 121 LVFGLTANSDTHLLOQOSLTLTLESPPGSSPSVQCHSPRGKNIQGGKTLVSQLELDQSG 180
 QY 181 TWTCVTLOQOKKVEFKIDIVLAFQKASSIVYKKEGEQVEFSPPLAFTYEKLTGSGELMW 240
 DB 181 TWTCVTLOQOKKVEFKIDIVLAFQKASSIVYKKEGEQVEFSPPLAFTYEKLTGSGELMW 240
 QY 241 QAEERASSSKSWITFDLKNKEVSVKRVTPQDPKLOMGKULPLHLTLPOLPOYAGSGNLTLLA 300
 DB 241 QAEERASSSKSWITFDLKNKEVSVKRVTPQDPKLOMGKULPLHLTLPOLPOYAGSGNLTLLA 300
 QY 301 LEAKTGKLGHEVNLVVMRAATQLOKNLTCEVWGPTSPKMLSLKLENKEAKVSRREKPVVW 360
 DB 301 LEAKTGKLGHEVNLVVMRAATQLOKNLTCEVWGPTSPKMLSLKLENKEAKVSRREKPVVW 360
 QY 361 LNPEAGMOCCLSDSGOVLLESNIKVLPTWSTPV 394
 DB 361 LNPEAGMOCCLSDSGOVLLESNIKVLPTWSTPV 394
 RESULT 64
 ID AAY88329 standard; protein; 458 AA.
 XX
 XX AAY88329;
 XX
 DT 14-JUL-2000 (first entry)
 XX
 XX T4 glycoprotein amino acid sequence.
 DE sT4; glycoprotein; human immunodeficiency virus; HIV; block binding;
 KM AIDS; treatment; inhibit; cell to cell spread; infection; fusion.

XX
 OS Mammalia.
 XX
 PN US5126433-A.
 XX
 PD 30-JUN-1992.
 XX
 XX 23-OCT-1987; 87US-00114244.
 PF
 PR 21-AUG-1986; 86US-00898587.
 XX
 XX (UYCO) UNIV COLUMBIA NEW YORK.
 PA
 PI Madden PJ, Chess L, Axel R, Weiss R, Littman DR, McDougal JS;
 XX MPI, 2000-348913/30.
 XX
 DR Soluble T4 glycoprotein useful for prevention and treatment of acquired
 PT immunodeficiency syndrome and for screening inhibitors of human
 PT immunodeficiency viral binding.
 XX
 XX Example; Fig 6; 64pp; English.
 XX
 CC This sequence represents the amino acid sequence of glycosylated sT4
 CC glycoprotein. Human immunodeficiency virus (HIV) uses sT4 as a target
 CC receptor on T cells. The invention relates to glycosylated sT4 which
 CC functions by blocking the binding of HIV to T4 target cells, and can be
 CC used for the prophylaxis and treatment of AIDS patients. Administration
 CC of sT4 effectively inhibits the cell to cell spreading of HIV infection
 CC and also the fusion of HIV-infected T4 cells and non-infected T4 cells.
 CC The administration of T4 alleviates several symptoms associated with
 CC AIDS, and prevents the occurrence of new pathological changes. The sT4
 CC glycoprotein is useful for the prophylaxis and treatment of patients with
 CC AIDS. It is also useful as a reagent to identify natural, synthetic or
 CC recombinant molecules which act as therapeutic agents or inhibitors of
 CC T4+ cell interactions and in diagnostic assays for detection T4 proteins
 CC or molecules
 CC
 XX
 SQ Sequence 458 AA;
 Query Match 59.0%; Score 2015; DB 3; Length 458;
 Best Local Similarity 99.5%; Pred. No. 6.8e-101;
 Matches 392; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MNRGVPFRHLILVQLALIPATQGNKVVGLKKGDVVELTCTASOKKSIQFHKNSNQIK 60
 DB 1 MNRGVPFRHLILVQLALIPATQGNKVVGLKKGDVVELTCTASOKKSIQFHKNSNQIK 60
 QY 61 ILGNQGSFLTKGPSKLNDRADRSRLMDQGNFPLIINKLKIEDSDTYICEVEDQKEEYVL 120
 DB 61 ILGNQGSFLTKGPSKLNDRADRSRLMDQGNFPLIINKLKIEDSDTYICEVEDQKEEYVL 120
 QY 121 LVFGLTANSDTHLLOQOSLTLTLESPPGSSPSVQCHSPRGKNIQGGKTLVSQLELDQSG 180
 DB 121 LVFGLTANSDTHLLOQOSLTLTLESPPGSSPSVQCHSPRGKNIQGGKTLVSQLELDQSG 180
 QY 181 TWTCVTLOQOKKVEFKIDIVLAFQKASSIVYKKEGEQVEFSPPLAFTYEKLTGSGELMW 240
 DB 181 TWTCVTLOQOKKVEFKIDIVLAFQKASSIVYKKEGEQVEFSPPLAFTYEKLTGSGELMW 240
 QY 241 QAEERASSSKSWITFDLKNKEVSVKRVTPQDPKLOMGKULPLHLTLPOLPOYAGSGNLTLLA 300
 DB 241 QAEERASSSKSWITFDLKNKEVSVKRVTPQDPKLOMGKULPLHLTLPOLPOYAGSGNLTLLA 300
 QY 301 LEAKTGKLGHEVNLVVMRAATQLOKNLTCEVWGPTSPKMLSLKLENKEAKVSRREKPVVW 360
 DB 301 LEAKTGKLGHEVNLVVMRAATQLOKNLTCEVWGPTSPKMLSLKLENKEAKVSRREKPVVW 360
 QY 361 LNPEAGMOCCLSDSGOVLLESNIKVLPTWSTPV 394
 DB 361 LNPEAGMOCCLSDSGOVLLESNIKVLPTWSTPV 394

RESULT 65
AAB81502
ID AAB81502 standard; protein: 458 AA.
XX
AC AAB81502;
XX
DT 18-JUN-2001 (first entry)
XX
DE Human CD4 protein.
XX
KM Human; CD4; CD4 fusion protein; oligomerisation;
KM receptor-ligand interaction inhibition; surface plasmon resonance; SPR;
KM T cell receptor binding; MHC binding; carcinoma; autoimmune disease;
KM multiple sclerosis; human immunodeficiency virus; HIV; diabetes;
KM rheumatoid arthritis; immune disorder.
XX
OS Homo sapiens.
XX
FH Key
FH Peptide 1. .25
FT /label= Signal_peptide
FT Protein 26. .458
FT /label= Human_CD4
XX
PN WO200122084-A2.
XX
PD 29-MAR-2001.
XX
PF 18-SEP-2000; 2000WO-GB003579.
XX
PR 21-SEP-1999; 99GB-00022352.
XX
PA (AVID-) AVIDEX LTD.
XX
PI Jakobsen BK;
XX
DR WPI; 2001-273470/28.
DR N-PSDB; AAF82582.
XX
PT Sequential screening of candidate compounds library for those which
PT binding of low affinity receptor-ligand interaction having fast
PT binding kinetics, using interfacial optical assay.
XX
PS Disclosure; Fig 13; 91pp; English.
XX
CC The present sequence is human CD4. Human CD4 extracellular domains 1 and
CC 2 were used in the construction of CD4 oligomerisation fusion proteins.
CC The fusion proteins contain an oligomerisation domain that enables the
CC proteins to bind to one another to form oligomers. The oligomers may be
CC used in an invention relating to a method for screening for compounds
CC with the ability to inhibit a low affinity receptor-ligand interaction.
CC The method uses an interfacial optical assay, such as surface plasmon
CC resonance (SPR). The method is useful for screening candidate compounds
CC for the ability to inhibit interaction between MHC/peptide complex and T
CC cell receptor, and MHC/peptide complex and CD8 or CD4 co-receptor. The
CC compounds identified by the above methods which interfere with T cell
CC receptor binding to a particular HLA type molecule are useful as immune
CC inhibitors for treating carcinomas, autoimmune diseases such as multiple
CC sclerosis, human immunodeficiency virus (HIV) infection, rheumatoid
CC arthritis, Hashimoto's disease, insulin dependent diabetes, Good
CC pasture's syndrome, uveitis, psoriasis and graft rejection
XX
SQ Sequence 458 AA;
XX
Query Match 59.0%; Score 2015; DB 4; Length 458;
Best Local Similarity 99.5%; Pred. No. 6.8e-101;
Matches 392; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 61 ILNGSGFLTKGPKLNDRADSRSLMDQGNPILIKNLKIEDSDTYICEVEDQKEEVL 120
Qy 121 LVFGLTANSDTHLLQGGSLTLTLESPGSSPSVQCSPRKNIQGGKTLTSSQLDELDSG 180
Db 121 LVFGLTANSDTHLLQGGSLTLTLESPGSSPSVQCSPRKNIQGGKTLTSSQLDELDSG 180
Qy 181 TWCTVYQONQKRYEFPKIDIVLAFQKASSIVYKKEGQVEFSPPLATVETLTSGLMW 240
Db 181 TWCTVYQONQKRYEFPKIDIVLAFQKASSIVYKKEGQVEFSPPLATVETLTSGLMW 240
Qy 241 QAERASSSKSWITFDLKNKEVSVKRVYQDPKLOMGKKLPALTLPOALPYAGSGNLTLA 300
Db 241 QAERASSSKSWITFDLKNKEVSVKRVYQDPKLOMGKKLPALTLPOALPYAGSGNLTLA 300
Qy 301 LEAKTGKLEHENVLVWRATQLOKNTLCEVWGPTSPKTLMLLENKAEVSKREKAVW 360
Db 301 LEAKTGKLEHENVLVWRATQLOKNTLCEVWGPTSPKTLMLLENKAEVSKREKAVW 360
Qy 361 LNPEAGMWOCILSDSGQVLESNIKVLPTWSTPY 394
Db 361 LNPEAGMWOCILSDSGQVLESNIKVLPTWSTPY 394
RESULT 66
ADD25609
ID ADD25609 standard; protein: 458 AA.
XX
AC ADD25609;
XX
DT 15-JAN-2004 (first entry)
XX
DE Binding domain-immunoglobulin fusion protein-associated protein #82.
XX
KM Binding domain; immunoglobulin; fusion protein; cytosolic;
KM antiarthritic; immunosuppressive; antidiabetic; antithyroid;
KM neuroprotective; hinge region; immunoglobulin heavy chain;
KM CH2 constant region; CH3 constant region; IgG1;
KM antibody dependent cell-mediated cytotoxicity; ADCC; complement fixation;
KM malignant condition; B-cell disorder; melanoma; carcinoma; sarcoma;
KM rheumatoid arthritis; myasthenia gravis; Grave's disease;
KM type I diabetes mellitus; multiple sclerosis; autoimmune disease.
XX
OS Unidentified.
XX
PN US2003118592-A1.
XX
PD 26-JUN-2003.
XX
PF 25-JUL-2002; 2002US-00207655.
XX
PR 17-JAN-2001; 2001US-0367358P.
PR 17-JAN-2002; 2002US-00053530.
PR 03-JUN-2002; 2002US-0385691P.
XX
PA (GENE-) GENE-CRAFT INC.
XX
PI Ledbetter JA, Hayden-Ledbetter MS, Thompson PA;
XX WPI; 2003-801317/75.
XX
PT New binding domain-immunoglobulin fusion protein, useful for treating a
PT subject having or suspected of having a malignant condition or a B-cell
PT disorder, e.g. melanoma, Grave's disease or autoimmune disease.
XX
PS Disclosure; SEQ ID NO 170; 157pp; English.
XX
CC The invention relates to a binding domain-immunoglobulin fusion protein
CC comprising a binding domain polypeptide that is fused to an
CC immunoglobulin hinge region polypeptide, an immunoglobulin heavy chain
CC CH2 constant region polypeptide that is fused to the hinge region
CC polypeptide, and an immunoglobulin heavy chain CH3 constant region
CC polypeptide that is fused to the CH2 constant region polypeptide. The

CC hinge region polypeptide comprises: a wild-type human IgG1 immunoglobulin hinge region polypeptide; a mutated human IgG1 immunoglobulin hinge region polypeptide, derived from (a) having 3 or more cysteine residues; where the mutated human IgG1 immunoglobulin hinge region polypeptide contains 2 cysteine residues, where the first cysteine is not mutated; a mutated human IgG1 immunoglobulin hinge region polypeptide, derived from (a) having 3 or more cysteine residues, where the mutated human IgG1 immunoglobulin hinge region polypeptide contains no more than one cysteine residue; and a mutated human IgG1 immunoglobulin hinge region polypeptide, derived from (a) having 3 or more cysteine residues; where the mutated human IgG1 immunoglobulin hinge region polypeptide contains no cysteine residues. The binding domain-immunoglobulin fusion protein is capable of at least one immunological activity comprising antibody dependent cell-mediated cytotoxicity (ADCC) and complement fixation. The binding domain polypeptide is capable of specifically binding to an antigen. Also included are an isolated polynucleotide encoding the binding domain-immunoglobulin fusion protein, a recombinant expression construct comprising the polynucleotide (operably linked to a promoter), a host cell transformed or transfected with a recombinant expression construct, producing the binding domain-immunoglobulin fusion protein, a pharmaceutical composition comprising the binding domain-immunoglobulin fusion protein or polynucleotide and a carrier, and treating a subject having or suspected of having a malignant condition or a B-cell disorder. The binding domain-immunoglobulin fusion protein is useful for treating a subject having or suspected of having a malignant condition or a B-cell disorder, e.g. melanoma, carcinoma or sarcoma, rheumatoid arthritis, myasthenia gravis, Grave's disease, type 1 diabetes mellitus, multiple sclerosis or autoimmune disease. The present sequence is a binding domain-immunoglobulin fusion protein-associated protein sequence. Note: The sequence data for this patent formed part of the printed specification and is also available in electronic format directly from USPTO at seqdata.uspto.gov/sequence.html?docid=20030118592. The authors have not identified the sequences in the printed specification by their SEQ ID number therefore none of the sequences can be explicitly identified.

SQ Sequence 458 AA;

Query Match 59.0%; Score 2015; DB 7; Length 458;

Best Local Similarity 99.5%; Pred. No. 6.8e-101;

Matches 392; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MNRGVFRRLLVLTQALLPAATQGNKVVVGKKGDVVELTCTASQKKSIOFHKNSNQIK 60
DB 1 MNRGVFRRLLVLTQALLPAATQGNKVVVGKKGDVVELTCTASQKKSIOFHKNSNQIK 60
QY 61 ILGNQGSFRTKPSKLNDRADSRSLMDQGNFLLIINKLIEDSDTYICEVEDQKEVOL 120
DB 61 ILGNQGSFRTKPSKLNDRADSRSLMDQGNFLLIINKLIEDSDTYICEVEDQKEVOL 120
QY 121 LVFGLTNSDTHLLQGGSLTLLTSPGSSPSVQCRRPKKNIOGGKTLVSQLELDQDSG 180
DB 121 LVFGLTNSDTHLLQGGSLTLLTSPGSSPSVQCRRPKKNIOGGKTLVSQLELDQDSG 180
QY 181 TWICTVQONQKVEFKIDIVLAFOKASSIVYKKEGQVFFSPLAFTVEKLTGSGELMW 240
DB 181 TWICTVQONQKVEFKIDIVLAFOKASSIVYKKEGQVFFSPLAFTVEKLTGSGELMW 240
QY 241 QABRASSKSWITFDLKNKEVSVRKVTQDPKLOMGKKLPHLTLPQALPOYAGSGNITLA 300
DB 241 QABRASSKSWITFDLKNKEVSVRKVTQDPKLOMGKKLPHLTLPQALPOYAGSGNITLA 300
QY 301 LEAKTGLHOEVNLVYWRATQLOKNTLCEVWGPTSPKLTMLKLENKAASVREKRWY 360
DB 301 LEAKTGLHOEVNLVYWRATQLOKNTLCEVWGPTSPKLTMLKLENKAASVREKRWY 360
QY 361 LNPEAGMWQCLISDSGOVLLESNIKVLPTWSTPV 394
DB 361 LNPEAGMWQCLISDSGOVLLESNIKVLPTWSTPV 394

RESULT 67
ADE57489
ID ADE57489 standard; protein: 458 AA.

XX ADE57489;
AC 29-JAN-2004 (first entry)
XX Human Protein P01730, SEQ ID NO 3351.
DE Human Protein P01730, SEQ ID NO 3351.
XX Human; pain; neuronal tissue; gene therapy;
KW spinal segmental nerve injury; chronic constriction injury; CCI;
XX spared nerve injury; SNI; Chung.
XX Homo sapiens.
XX MO2003016475-A2.
XX 27-FEB-2003.
XX 14-AUG-2002; 2002WO-US025765.
XX 14-AUG-2001; 2001US-0312147P.
PR 01-NOV-2001; 2001US-0346382P.
PR 26-NOV-2001; 2001US-033347P.
XX (GENO) GEN HOSPITAL CORP.
PA (FAR) BAYER AG.
XX Woolf C, D'urso D, Befort K, Costigan M;
XX WPI; 2003-268312/26.
DR GENBANK; P01730.
XX New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.
XX Claim 1; Page: 1017pp; English.
PS The invention discloses a composition comprising two or more isolated rat
XX or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a human protein (shown in Table 2 of
CC the specification) which is differentially expressed during pain. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX
SQ Sequence 458 AA;

Query Match 59.0%; Score 2015; DB 7; Length 458;

Best Local Similarity 99.5%; Pred. No. 6.8e-101;

Matches 392; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MNRGVFRRLLVLTQALLPAATQGNKVVVGKKGDVVELTCTASQKKSIOFHKNSNQIK 60
DB 1 MNRGVFRRLLVLTQALLPAATQGNKVVVGKKGDVVELTCTASQKKSIOFHKNSNQIK 60

QY ILGNQGSFLTKGPKLNDRADSRSLMDQGNFPLIKNLKIEDSDTYICEVEDQKEEVOL 120
 DB 61 ILGNQGSFLTKGPKLNDRADSRSLMDQGNFPLIKNLKIEDSDTYICEVEDQKEEVOL 120
 QY 121 LVFGLTANSDTHLLQGGSLTLTLESPPGSSPSVQCRSPRKNIOGKTLVSQLELQDSG 180
 DB 121 LVFGLTANSDTHLLQGGSLTLTLESPPGSSPSVQCRSPRKNIOGKTLVSQLELQDSG 180
 QY 181 TWCTVQONQKVEFKIDIVVLAFOKASSIYKKEGEQVEVSFPLAFTVEKLTGSGELMW 240
 DB 181 TWCTVQONQKVEFKIDIVVLAFOKASSIYKKEGEQVEVSFPLAFTVEKLTGSGELMW 240
 QY 241 QAERASSSSKSWITFDLKNKEVSVRVYODPKLQMGKLPPLHTLPOALPOYAGSGNLTIA 300
 DB 241 QAERASSSSKSWITFDLKNKEVSVRVYODPKLQMGKLPPLHTLPOALPOYAGSGNLTIA 300
 QY 301 LEAKTGKLGHEVNLVVMRATOLQKNLTCEVWGFTSPKLMSTLKLLENKAKVSKREKAPVW 360
 DB 301 LEAKTGKLGHEVNLVVMRATOLQKNLTCEVWGFTSPKLMSTLKLLENKAKVSKREKAPVW 360
 QY 361 LNPEAGMWQCLLSDSGVLLSNIKVLPTWSTPV 394
 DB 361 LNPEAGMWQCLLSDSGVLLSNIKVLPTWSTPV 394

RESULT 68

ADA44807
 ID ADA44807 standard; protein; 473 AA.

AC ADA44807;

DT 04-DEC-2003 (first entry)

DE CD4/TCR CD3epsilon chain chimeric protein CD4epsilon10n15, SEQ ID NO:2.

KM HIV-1 infection; human immunodeficiency virus-1; CD4+ cell; chimeric CD4;

KM endoplasmic reticulum; ER retention; envelope protein gp160;

KM T cell receptor CD3epsilon chain; C-terminal domain; CD4epsilon10n15;

XX gene therapy; human; receptor.

OS Chimeric.

XX Homo sapiens.

XX Key

XX Protein

XX Region

XX 459..473

XX receptor CD3epsilon chain"

XX WO2003076468-A1.

XX 18-SEP-2003.

XX 14-MAR-2003; 2003WO-ES000120.

XX 14-MAR-2002; 2002ES-0000616.

XX (NSJ) CONSEJO SUPERIOR INVESTIGACIONES CIENTIF.

XX Alarcon Sanchez BJ, San Jose Martinez ME, Zaldivar Nocardio I;

XX Gomez Buendia M;

XX WPI; 2003-779059/73.

XX N-PSDB; ADA44806.

XX Composition for treating or preventing human immune deficiency virus,

XX comprises CD4 chimeric protein having a protective effect in trans, or

XX related nucleic acid.

CC The invention relates to a composition for the treatment or prevention of
 CC human immunodeficiency virus-1 (HIV-1) infection. The composition
 CC comprises CD4+ cells that have been transduced with a vector that encodes
 CC a chimeric CD4 molecule which is capable of being retained in the
 CC endoplasmic reticulum (ER). The invention also encompasses the use of a
 CC soluble protein factor produced by CD4+ cells that have been transduced
 CC with a vector encoding a chimeric CD4 protein; and the use of an
 CC expression system encoding a chimeric CD4 protein. The ER-located
 CC chimeric CD4 molecule binds to the HIV-1 envelope protein gp160,
 CC resulting in HIV-1 retention in the ER and thereby preventing viral
 CC replication. In a specific embodiment, the chimeric CD4 molecule
 CC comprises CD4 fused to 15 amino acids of the C-terminal domain of the T
 CC cell receptor CD3epsilon chain; this chimeric CD4 molecule is designated
 CC CD4epsilon10n15 (ADA44807). A known chimeric CD4 of similar structure but
 CC containing only 10 amino acids from CD3epsilon can also be used.
 CC Compositions of the invention have an in trans effect on the replication
 CC of HIV-1, and may be used to treat and prevent HIV-1 infection. The
 CC present sequence represents the chimeric CD4 molecule CD4epsilon10n15, which
 CC is specifically claimed for use in compositions of the invention.
 XX Sequence 473 AA;
 SQ
 QY 1 MNRGVPPRHLLVYQLALPAAATQGNKVYLGKGDVYELTCTASQKKSIOFHWKNSQIK 60
 DB 1 MNRGVPPRHLLVYQLALPAAATQGNKVYLGKGDVYELTCTASQKKSIOFHWKNSQIK 60
 QY 61 ILGNQGSFLTKGPKLNDRADSRSLMDQGNFPLIKNLKIEDSDTYICEVEDQKEEVOL 120
 DB 61 ILGNQGSFLTKGPKLNDRADSRSLMDQGNFPLIKNLKIEDSDTYICEVEDQKEEVOL 120
 QY 121 LVFGLTANSDTHLLQGGSLTLTLESPPGSSPSVQCRSPRKNIOGKTLVSQLELQDSG 180
 DB 121 LVFGLTANSDTHLLQGGSLTLTLESPPGSSPSVQCRSPRKNIOGKTLVSQLELQDSG 180
 QY 181 TWCTVQONQKVEFKIDIVVLAFOKASSIYKKEGEQVEVSFPLAFTVEKLTGSGELMW 240
 DB 181 TWCTVQONQKVEFKIDIVVLAFOKASSIYKKEGEQVEVSFPLAFTVEKLTGSGELMW 240
 QY 241 QAERASSSSKSWITFDLKNKEVSVRVYODPKLQMGKLPPLHTLPOALPOYAGSGNLTIA 300
 DB 241 QAERASSSSKSWITFDLKNKEVSVRVYODPKLQMGKLPPLHTLPOALPOYAGSGNLTIA 300
 QY 301 LEAKTGKLGHEVNLVVMRATOLQKNLTCEVWGFTSPKLMSTLKLLENKAKVSKREKAPVW 360
 DB 301 LEAKTGKLGHEVNLVVMRATOLQKNLTCEVWGFTSPKLMSTLKLLENKAKVSKREKAPVW 360
 QY 361 LNPEAGMWQCLLSDSGVLLSNIKVLPTWSTPV 394
 DB 361 LNPEAGMWQCLLSDSGVLLSNIKVLPTWSTPV 394

RESULT 69

AAR20152
 ID AAR20152 standard; protein; 519 AA.

XX AAR20152;

XX 25-MAR-2003 (revised)

XX 31-MAR-1992 (first entry)

XX Human CD4 sequence encoded by PATY. 6.

XX Human immunodeficiency virus; HIV; gp 120; AIDS; ARC; glycoprotein;

XX acquired immune deficiency syndrome; AIDS related complex;

XX T helper lymphocytes.

XX Homo sapiens.

XX Key

XX Location/Qualifiers

```

FT Peptide 1..25
FT /label= signal_sequence
XX PN WO9118618-A.
XX PD 12-DEC-1991.
XX PF 25-MAY-1990; 90US-00529186.
XX PR 25-MAY-1990; 90US-00529186.
XX PA (BIOU ) BIOGEN INC.
XX PI Fisher RA, Hession C, Burkly LC;
XX DR WPI, 1992-007200/01.
XX DR N-PSDB; AAQ20327.
XX PT New immuno-therapeutic human CD4 variants and deriva. - elicit AB
XX PT production to HIV gp120, useful in treating, preventing and diagnosing
XX PT AIDS, ARC and HIV infections.
XX PS Disclosure; Fig 28; 179pp; English.
XX CC The sequence was deduced from the DNA sequence of subclone PATY.6, contg.
XX CC DNA coding for the full-length human CD4. The clone was constructed from
XX CC plasmids PB6178A and PB6378 (both in US8802940). The DNA can be used to
XX CC express recombinant CD4 and analogues for use in diagnosis and treatment
XX CC of diseases caused by infective agents whose primary targets are T4+
XX CC lymphocytes. See also AAR20148-R20155 and AAR21078. (Updated on 25-MAR-
XX CC 2003 to correct PA field.)
SQ Sequence 519 AA;

Query Match 59.0%; Score 2015; DB 2; Length 519;
Best Local Similarity 99.5%; Pred. No. 7.8e-101;
Matches 392; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MNRGVFRRHLVLVQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIOFHKNSNOIK 60
DB 62 MNRGVFRRHLVLVQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIOFHKNSNOIK 121
QY 61 ILGNQGSFLTKGSPSKLNDRAISRSLMDQGNFPLIINKLKIENSDTYICVEVDQKEVOL 120
DB 122 ILGNQGSFLTKGSPSKLNDRAISRSLMDQGNFPLIINKLKIENSDTYICVEVDQKEVOL 181
QY 121 LVFGLTANSPTHLLOQGSLLTLESPPGSSPSVQCRSPRGKNIQGGKTLISVQLELDQSG 180
DB 182 LVFGLTANSPTHLLOQGSLLTLESPPGSSPSVQCRSPRGKNIQGGKTLISVQLELDQSG 241
QY 181 TWICTTVLQONKQVEFKIDIVVLAFOKASSIVYKKEGQVFEFPLATVTEKLTGSGELMW 240
DB 242 TWICTTVLQONKQVEFKIDIVVLAFOKASSIVYKKEGQVFEFPLATVTEKLTGSGELMW 301
QY 241 QAERASSSKSMITFDLKNKEVSVKRVTPDPLQMGKTLPHLTLPOLPOYASSGNLTLLA 300
DB 302 QAERASSSKSMITFDLKNKEVSVKRVTPDPLQMGKTLPHLTLPOLPOYASSGNLTLLA 361
QY 301 LEAKTGLHQBVLVVMRATQLOKNTLCEVWGPTSPKMLSLKLENKAVSKREKRVW 360
DB 362 LEAKTGLHQBVLVVMRATQLOKNTLCEVWGPTSPKMLSLKLENKAVSKREKRVW 421
QY 361 LNPEAGMWQCLISDSGVLLSNIKVLPWTSTPV 364
DB 422 LNPEAGMWQCLISDSGVLLSNIKVLPWTSTPV 455

```

RESULT 70

```

AAR13491
ID AAR13491 standard; protein; 458 AA.
XX AC AAR13491;
XX

```

```

DT 25-MAR-2003 (revised)
DT 30-OCT-1991 (first entry)
XX DE Human CD4 encoded by pUD.SCD4.Y187.Snab1 and p170.2.
XX DE C4bp; gp120; HIV; T lymphocyte; fusion protein.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT Peptide 1..25
XX FT /label= signal_peptide
XX FT Domain 25..132
XX FT /label= Ig-related
XX FT /note= "extracellular"
XX FT Disulfide-bond 41..109
XX FT Domain 133..202
XX FT /label= Ig-related
XX FT /note= "extracellular"
XX FT Disulfide-bond 155..184
XX FT Domain 203..318
XX FT /label= Ig-related
XX FT /note= "extracellular"
XX FT Domain 319..395
XX FT /label= Ig-related
XX FT /note= "extracellular"
XX FT Disulfide-bond 328..370
XX FT Region 356..416
XX FT /label= transmembrane
XX FT Domain 417..456
XX FT /label= cytoplasmic

PN WO9111461-A.
XX PD 08-AUG-1991.
XX PF 26-JAN-1990; 90US-00470888.
XX PR 26-JAN-1990; 90US-00470888.
XX PA (BIOU ) BIOGEN INC.
XX PI Pasek MP, Winkler G, Liu TR;
XX DR WPI, 1991-252613/34.
XX DR N-PSDB; AAQ13243.
XX PT New C4 binding protein fusion proteins and DNA encoding them - comprise
XX PT assemblies of C4bp monomers linked to functional moiety, e.g. AZT, useful
XX PT as delivery vehicles in diagnosis and therapy.
XX PS Example 3; Fig 3; 105pp; English.
XX CC This is the preferred CD4 sequence for use in the construction of fusion
XX CC proteins with C4-binding protein. Truncated, soluble versions of CD4 can
XX CC also be used. The C4bp-CD4 fusion protein may be useful to target AZT or
XX CC similar anti-retroviral agent to HIV-infected cells. See AAQ13242-51.
XX CC (Updated on 25-MAR-2003 to correct PA field.)
SQ Sequence 458 AA;

Query Match 59.0%; Score 2013; DB 2; Length 458;
Best Local Similarity 99.5%; Pred. No. 8.8e-101;
Matches 392; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MNRGVFRRHLVLVQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIOFHKNSNOIK 60
DB 62 MNRGVFRRHLVLVQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIOFHKNSNOIK 121
QY 61 ILGNQGSFLTKGSPSKLNDRAISRSLMDQGNFPLIINKLKIENSDTYICVEVDQKEVOL 120
DB 61 ILGNQGSFLTKGSPSKLNDRAISRSLMDQGNFPLIINKLKIENSDTYICVEVDQKEVOL 120

```

QY 121 LVFGLTANSDTHLLQGSLTTLTLESPPGSSPSVOCSPRGKNIQGGKTLVSQLELDQSG 180
 DB 121 LVFGLTANSDTHLLQGSLTTLTLESPPGSSPSVOCSPRGKNIQGGKTLVSQLELDQSG 180
 QY 181 TWTCCTVLOQOKKVEPKIDIVLAFQKASSIYKKEGEQVEFSPPLAFTVEKLTGSGELMW 240
 DB 181 TWTCCTVLOQOKKVEPKIDIVLAFQKASSIYKKEGEQVEFSPPLAFTVEKLTGSGELMW 240
 QY 241 QAERASSKSWITFDLKNKEVSVRVYTOPDKLQMGKKLPLHLTLPOLPOYASGSLTLA 300
 DB 241 QAERASSKSWITFDLKNKEVSVRVYTOPDKLQMGKKLPLHLTLPOLPOYASGSLTLA 300
 QY 301 LEAKTGKLGHEVNLVWMRATQLOKNLTCEVWGPTSPKMLSLKLENKEAKVSKREKPEVMV 360
 DB 301 LEAKTGKLGHEVNLVWMRATQLOKNLTCEVWGPTSPKMLSLKLENKEAKVSKREKPEVMV 360
 QY 361 LNPEAGMWQCLLSDSGVLLSNIKVLPTWSTPV 394
 DB 361 LNPEAGMWQCLLSDSGVLLSNIKVLPTWSTPV 394

RESULT 71

AAI39825
 ID AAY39825 standard; protein; 394 AA.
 AC AAY39825;
 XX
 DT 03-DEC-1999 (first entry)
 XX
 DE Soluble human T4 protein.
 XX
 KM Soluble T4 protein; sT4; human; HIV; binding inhibitor; T4+ cell; AIDS;
 KW vaccine; immunisation; therapy.
 XX
 OS Homo sapiens.
 XX
 PN US958678-A.
 XX
 PD 28-SEP-1999.
 XX
 PF 12-DEC-1994; 94US-00354452.
 XX
 PR 21-AUG-1986; 86US-00898587.
 PR 11-JUN-1991; 91US-00713564.
 PR 06-JUL-1992; 92US-00909021.
 XX
 PA (UYCO) UNIV COLUMBIA NEW YORK.
 PI
 PI McDougal JS, Weiss R, Axel R, Littman DR, Maddon PJ, Chess L;
 DR WPI; 1999-561025/47.
 DR N-PSDB; AAZ20694.
 XX
 PT Human T4 protein inhibits HIV binding to T4 cells, useful for treating
 PT AIDS.
 XX
 PS Disclosure; Col 13-16; 58pp; English.
 XX
 CC This sequence represents the soluble human T4 protein of the invention.
 CC The soluble human T4 protein blocks the binding of HIV to T4+ cells and
 CC is therefore useful for the treatment of AIDS. Monoclonal antibodies
 CC against the T4 protein may be used as vaccines for immunising subjects
 CC against AIDS
 CC
 XX Sequence 394 AA;
 SQ

Query Match 58.9%; Score 1012; DB 2; Length 394;
 Best Local Similarity 99.2%; Pred. No. 8.5e-101;
 Matches 391; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MNRCVPRHLLLVQLALLPAATGKNVVLGKKGDVETLCTASQKSIQFHWKNSNQIK 60
 DB 1 MNRCVPRHLLLVQLALLPAATGKNVVLGKKGDVETLCTASQKSIQFHWKNSNQIK 60

QY 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQNPFLITKNLKIENSDPYICEVENOKKEVOL 120
 DB 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQNPFLITKNLKIENSDPYICEVENOKKEVOL 120
 QY 121 LVFGLTANSDTHLLQGSLTTLTLESPPGSSPSVOCSPRGKNIQGGKTLVSQLELDQSG 180
 DB 121 LVFGLTANSDTHLLQGSLTTLTLESPPGSSPSVOCSPRGKNIQGGKTLVSQLELDQSG 180
 QY 181 TWTCCTVLOQOKKVEPKIDIVLAFQKASSIYKKEGEQVEFSPPLAFTVEKLTGSGELMW 240
 DB 181 TWTCCTVLOQOKKVEPKIDIVLAFQKASSIYKKEGEQVEFSPPLAFTVEKLTGSGELMW 240
 QY 241 QAERASSKSWITFDLKNKEVSVRVYTOPDKLQMGKKLPLHLTLPOLPOYASGSLTLA 300
 DB 241 QAERASSKSWITFDLKNKEVSVRVYTOPDKLQMGKKLPLHLTLPOLPOYASGSLTLA 300
 QY 301 LEAKTGKLGHEVNLVWMRATQLOKNLTCEVWGPTSPKMLSLKLENKEAKVSKREKPEVMV 360
 DB 301 LEAKTGKLGHEVNLVWMRATQLOKNLTCEVWGPTSPKMLSLKLENKEAKVSKREKPEVMV 360
 QY 361 LNPEAGMWQCLLSDSGVLLSNIKVLPTWSTPV 394
 DB 361 LNPEAGMWQCLLSDSGVLLSNIKVLPTWSTPV 394

RESULT 72

AAG79087
 ID AAG79087 standard; protein; 458 AA.
 AC AAG79087;
 XX
 DT 10-DEC-2001 (first entry)
 XX
 DE Amino acid sequence of human CD4 protein.
 XX
 KM Human; receptor; DC-SIGN; dendritic cell; T lymphocyte; HIV; gp120;
 KW C-type lectin; ICM3; HIV entry; T cell; macrophage; HIV infection; CD4.
 XX
 OS Homo sapiens.
 XX
 PN WO200164752-A2.
 XX
 PD 07-SEP-2001.
 XX
 PF 28-FEB-2001; 2001WO-US006322.
 PF 02-MAR-2000; 2000US-00517605.
 PR
 PR (UTNY) UNIV NEW YORK STATE.
 PA (UTNY-) UNIV NIJMEGEN.
 XX
 PI Littman DR, Kwon D, Van Kooyk Y, Geijtenbeek T;
 PI WPI; 2001-602565/68.
 DR
 XX
 XX An antibody for the treatment or prevention of HIV-infection comprises a
 PT gp120 portion which binds to DC-SIGN or is exposed upon gp120 binding of
 PT DC-SIGN due to concomitant conformational change.
 XX
 PS Disclosure; Page 115-116; 131pp; English.
 XX
 CC The specification describes an antibody which is specific for an
 CC antigenic fragment of gp120. This antigenic fragment binds to DC-SIGN or
 CC is exposed upon gp120 binding of DC-SIGN due to concomitant
 CC conformational change. DC-SIGN is a receptor that is specifically
 CC expressed on dendritic cells and facilitates infection of T lymphocytes
 CC with HIV. DC-SIGN is identical to a HIV-1 gp120-binding C-type lectin. DC
 CC -SIGN binds ICM-3 (which is expressed constitutively on T lymphocytes)
 CC with high affinity. The antibody of the invention inhibits the trans
 CC enhancement of HIV entry into a T cell or macrophage facilitated by
 CC dendritic cells. The antibody is useful to treat or prevent HIV
 CC infection. The present sequence represents a human CD4 protein

```

XX
SQ Sequence 458 AA;
Query Match 58.8%; Score 2007; DB 4; Length 458;
Best Local Similarity 99.2%; Pred. No. 1.8e-100;
Matches 391; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MNRGVPFRHLVLVQLALIPATQGNKVVYLGKKGDTVELCTASQKKSIOFHWKNSNQIK 60
DB 1 MNRGVPFRHLVLVQLALIPATQGNKVVYLGKKGDTVELCTASQKKSIOFHWKNSNQIK 60
QY 61 ILGNQSSFLTKGPSKLNDRADSRSLMDQGNPFLIIKNLIKEDSDTYICEVEDQKEEYOL 120
DB 61 ILGNQSSFLTKGPSKLNDRADSRSLMDQGNPFLIIKNLIKEDSDTYICEVEDQKEEYOL 120
QY 121 LVFGLTANSDTHLLQGGSLTLTLSPGSSPSVQCRSPRGKNIQGGKTLVSQLEIQDSG 180
DB 121 LVFGLTANSDTHLLQGGSLTLTLSPGSSPSVQCRSPRGKNIQGGKTLVSQLEIQDSG 180
QY 181 TWCTVLOQNKQKVEFKIDIVLAFOKASSIVYKKEGEQVEFSPLAFTVEKLTGSGELMW 240
DB 181 TWCTVLOQNKQKVEFKIDIVLAFOKASSIVYKKEGEQVEFSPLAFTVEKLTGSGELMW 240
QY 241 QAERASSSKSWITFDLKNKEVSVKRYTODPKLOMGKKLPJHLTLPQALPOYAGSGNLTIA 300
DB 241 QAERASSSKSWITFDLKNKEVSVKRYTODPKLOMGKKLPJHLTLPQALPOYAGSGNLTIA 300
QY 301 LEAKTGKLEHGVNLVVMRATQLOKNLTCEVWGPTSPKMLSLKENKEAKVSKREKAVVW 360
DB 301 LEAKTGKLEHGVNLVVMRATQLOKNLTCEVWGPTSPKMLSLKENKEAKVSKREKAVVW 360
QY 361 LNPEAGMOCCLSDSGOVLLESNIKVLPTWSTPV 394
DB 361 LNPEAGMOCCLSDSGOVLLESNIKVLPTWSTPV 394

RESULT 73
AAB07768
ID AAB07768 standard; protein; 394 AA.
XX
AC AAB07768;
XX
DT 07-NOV-2000 (first entry)
XX
DE The soluble extracellular domain of the T4 glycoprotein.
XX
KW Human; T4 glycoprotein; human immunodeficiency virus; HIV;
KM envelope glycoprotein; AIDS; virus binding.
XX
OS Homo sapiens.
XX
PN US6093539-A.
XX
PD 25-JUL-2000.
XX
PF 06-JUN-1995; 95US-00466368.
XX
PR 21-AUG-1986; 86US-00898587.
PR 11-JUN-1991; 91US-00713564.
PR 06-JUL-1992; 92US-00909021.
PR 12-DEC-1994; 94US-00354452.
XX
PA (UYCO ) UNIV COLUMBIA NEW YORK.
XX
PI Madden PJ, Chess L, Axel R, Weiss R, McDougal JS, Littman DR,
XX
DR WPI; 2000-505203/45.
XX
DR N-PSDB; AAA59351.
XX
PT New isolated nucleic acid encoding a human T cell surface protein and the
PT soluble surface T4 glycoprotein that it encodes, useful as prophylaxis
PT for treating a subject infected with human acquired immune deficiency
PT syndrome virus.

```

```

XX
PS Disclosure; Col 11-14; 69pp; English.
XX
CC The present sequence represents an aqueous-soluble polypeptide comprising
CC a portion of a human T4 glycoprotein. The portion specifically forms a
CC complex with a human immunodeficiency virus (HIV) envelope glycoprotein.
CC The DNA is useful for producing the soluble surface T4 glycoprotein. The
CC soluble surface T4 glycoprotein is useful as a therapeutic agent, i.e. as
CC prophylaxis for treating a subject infected with an HIV virus. Thus, the
CC soluble T4 glycoprotein is useful for treating human AIDS. The soluble T4
CC glycoprotein is also useful in diagnostic or screening assays, e.g. for
CC screening inhibitors of virus binding, or for detecting and quantitating
CC T4, T4+ cells and antibodies to T4, which are of diagnostic value for
CC AIDS
XX
SQ Sequence 394 AA;
Query Match 58.6%; Score 2001; DB 3; Length 394;
Best Local Similarity 98.7%; Pred. No. 3.3e-100;
Matches 389; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 MNRGVPFRHLVLVQLALIPATQGNKVVYLGKKGDTVELCTASQKKSIOFHWKNSNQIK 60
DB 1 MNRGVPFRHLVLVQLALIPATQGNKVVYLGKKGDTVELCTASQKKSIOFHWKNSNQIK 60
QY 61 ILGNQSSFLTKGPSKLNDRADSRSLMDQGNPFLIIKNLIKEDSDTYICEVEDQKEEYOL 120
DB 61 ILGNQSSFLTKGPSKLNDRADSRSLMDQGNPFLIIKNLIKEDSDTYICEVEDQKEEYOL 120
QY 121 LVFGLTANSDTHLLQGGSLTLTLSPGSSPSVQCRSPRGKNIQGGKTLVSQLEIQDSG 180
DB 121 LVFGLTANSDTHLLQGGSLTLTLSPGSSPSVQCRSPRGKNIQGGKTLVSQLEIQDSG 180
QY 181 TWCTVLOQNKQKVEFKIDIVLAFOKASSIVYKKEGEQVEFSPLAFTVEKLTGSGELMW 240
DB 181 TWCTVLOQNKQKVEFKIDIVLAFOKASSIVYKKEGEQVEFSPLAFTVEKLTGSGELMW 240
QY 241 QAERASSSKSWITFDLKNKEVSVKRYTODPKLOMGKKLPJHLTLPQALPOYAGSGNLTIA 300
DB 241 QAERASSSKSWITFDLKNKEVSVKRYTODPKLOMGKKLPJHLTLPQALPOYAGSGNLTIA 300
QY 301 LEAKTGKLEHGVNLVVMRATQLOKNLTCEVWGPTSPKMLSLKENKEAKVSKREKAVVW 360
DB 301 LEAKTGKLEHGVNLVVMRATQLOKNLTCEVWGPTSPKMLSLKENKEAKVSKREKAVVW 360
QY 361 LNPEAGMOCCLSDSGOVLLESNIKVLPTWSTPV 394
DB 361 LNPEAGMOCCLSDSGOVLLESNIKVLPTWSTPV 394

RESULT 74
ADE65841
ID ADE65841 standard; protein; 458 AA.
XX
AC ADE65841;
XX
DT 29-JAN-2004 (first entry)
XX
DE Human CD4 receptor.
XX
KW Human; CD4 receptor; receptor; protein-protein interaction;
KM protein array; PDZ domain; drug target screening.
XX
OS Homo sapiens.
XX
PN US2003170723-A1.
XX
PD 11-SEP-2003.
XX
PF 06-MAR-2002; 2002US-00092138.
XX
PR 06-MAR-2002; 2002US-00092138.
XX

```

PA (SATO/) SATO T.
 XX
 XX
 PI Sato T;
 XX
 XX WPI: 2003-852032/79.
 XX
 PT Preparing a protein array useful for screening drug targets comprises
 PT depositing an array of a first protein on substrate, and applying a
 PT second protein comprising an amino acid sequence that binds to a domain
 PT of the first protein.
 XX
 PS Disclosure; SEQ ID NO 25; 60pp; English.
 XX
 CC The invention relates to a method for preparing a protein array based on
 CC protein-protein interaction, by depositing an array of a first protein
 CC comprising a PDZ domain on a substrate, and applying a second protein
 CC comprising an amino acid sequence that binds to the PDZ domain of the
 CC first protein. The method is useful for preparing protein arrays based on
 CC biochemical protein-protein interactions. Arrays produced by this method
 CC are useful for screening drug targets. This sequence represents the human
 CC CD4 receptor, used in the method of the invention.
 XX
 SQ Sequence 458 AA;
 Query Match 58.5%; Score 1997; DB 7; Length 458;
 Best Local Similarity 98.7%; Pred. No. 6.4e-100;
 Matches 389; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 QY 1 MNRGVPRHLLVLTQALLPAATQGNKVVLGKGGDTVELTCTASQKKSIOFHMKNNSQIK 60
 DB 1 MNRGVPRHLLVLTQALLPAATQGNKVVLGKGGDTVELTCTASQKKSIOFHMKNNSQIK 60
 QY 61 ILGNQGSFLTKGSPSKLNDRADSRSLMDQGNPILIKNKIEDSDTYICEVEDQKEVQL 120
 DB 61 ILGNQGSFLTKGSPSKLNDRADSRSLMDQGNPILIKNKIEDSDTYICEVEDQKEVQL 120
 QY 121 LVFGLTANSDTHLLQGGSLTTLTSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
 DB 121 LVFGLTANSDTHLLQGGSLTTLTSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
 QY 181 TWCTCTVLQNGKVEFKIDIVLAFQKASSIYKKEGEQVFSPLAFTVEKLTGSGELMW 240
 DB 181 TWCTCTVLQNGKVEFKIDIVLAFQKASSIYKKEGEQVFSPLAFTVEKLTGSGELMW 240
 QY 241 QAERASSSKSWITFDLKNKEVSVKRVYODPKLQMGKKLPHLTLPLQALPOYAGSGNLTIA 300
 DB 241 QAERASSSKSWITFDLKNKEVSVKRVYODPKLQMGKKLPHLTLPLQALPOYAGSGNLTIA 300
 QY 301 LEATGKGLHQEVNLVVMRATOLQKNTCEVWGPTSPKMLSLKLENKEAKVSKKEKPYMV 360
 DB 301 LEATGKGLHQEVNLVVMRATOLQKNTCEVWGPTSPKMLSLKLENKEAKVSKKEKPYMV 360
 QY 361 LNPAGMWQCLLSDSGQVLTLESNIKVLPTWSTPV 394
 DB 361 LNPAGMWQCLLSDSGQVLTLESNIKVLPTWSTPV 394
 RESULT 75
 AAR06373
 ID AAR06373 standard; protein; 458 AA.
 XX
 AC AAR06373;
 XX
 DT 31-OCT-2002 (revised)
 DT 20-DEC-1990 (first entry)
 XX
 XX T4 encoded by plasmid p170-2.
 DE
 XX plasmid p170-2; soluble T4 protein; AIDS; ARC; HIV.
 XX
 OS Synthetic.
 XX
 XX Key Location/Qualifiers

FT Peptide
 FT 1. 23
 FT /label= hydrophobic/secretory signal
 FT Region
 FT 24. 117
 FT /label= extracellular
 FT /note= "homology to V-regions"
 FT Region
 FT 118. 132
 FT /label= extracellular
 FT /note= "homology to J-regions"
 FT Region
 FT 133. 397
 FT /label= extracellular
 FT /note= "glycosylated region"
 FT Region
 FT 398. 418
 FT /label= transmembrane sequence
 FT /note= "hydrophobic"
 FT Region
 FT 419. 458
 FT /label= intracytoplasmic
 FT /note= "very hydrophilic"
 PN W09008198-A.
 XX 26-JUL-1990.
 PD
 XX 18-JAN-1989; 89US-00300096.
 PF
 XX 18-JAN-1989; 89US-00300096.
 PR
 XX (HARD) HARVARD COLLEGE.
 PA
 XX Letvin NA;
 PI
 XX WPI: 1990-254040/33.
 DR N-PSDB; AAQ05607.
 XX
 PT Treating or preventing AIDS, ARC or HIV infection - by administering an
 PT immunologically effective amt. of soluble T4 protein.
 PT
 XX
 PS Disclosure; Fig 1; 121pp; English.
 CC Soluble T4 can be produced by truncating the CDS to remove the
 CC transmembrane and cytoplasmic domains. The soluble forms may be modified
 CC to increase their immunogenicity by addition of an adjuvant such as
 CC incomplete Freund's adjuvant. The T4 interferes with HIV/T4 interaction
 CC and elicits anti-soluble T4 antibody production. Soluble T4 proteins
 CC include the following polypeptide fragments: amino acids 1-385, 24-385,
 CC Met-24-385, 24-397, 1-400 and Met-24-400. See also AAQ05606. (Updated on
 CC 31-OCT-2002 to add missing OS field.)
 CC
 SQ Sequence 458 AA;
 Query Match 58.4%; Score 1993; DB 2; Length 458;
 Best Local Similarity 99.0%; Pred. No. 1e-99;
 Matches 390; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 MNRGVPRHLLVLTQALLPAATQGNKVVLGKGGDTVELTCTASQKKSIOFHMKNNSQIK 60
 DB 1 MNRGVPRHLLVLTQALLPAATQGNKVVLGKGGDTVELTCTASQKKSIOFHMKNNSQIK 60
 QY 61 ILGNQGSFLTKGSPSKLNDRADSRSLMDQGNPILIKNKIEDSDTYICEVEDQKEVQL 120
 DB 61 ILGNQGSFLTKGSPSKLNDRADSRSLMDQGNPILIKNKIEDSDTYICEVEDQKEVQL 120
 QY 121 LVFGLTANSDTHLLQGGSLTTLTSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
 DB 121 LVFGLTANSDTHLLQGGSLTTLTSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
 QY 181 TWCTCTVLQNGKVEFKIDIVLAFQKASSIYKKEGEQVFSPLAFTVEKLTGSGELMW 240
 DB 181 TWCTCTVLQNGKVEFKIDIVLAFQKASSIYKKEGEQVFSPLAFTVEKLTGSGELMW 240
 QY 241 QAERASSSKSWITFDLKNKEVSVKRVYODPKLQMGKKLPHLTLPLQALPOYAGSGNLTIA 300
 DB 241 QAERASSSKSWITFDLKNKEVSVKRVYODPKLQMGKKLPHLTLPLQALPOYAGSGNLTIA 300

```

QY 301 LEAKTGKTLHQEVNLYVMRATOLQKNLTCEVWGPTSPKMLSLKLNKKAQVSKREKPVV 360
DB 301 LEAKTGKTLHQEVNLYVMRATOLQKNLTCEVWGPTSPKMLSLKLNKKAQVSKREKPVV 360
QY 361 LNPEAGMOCCLSDSGQVLLSENIKYLPTWSTPV 394
DB 361 LNPEAGMOCCLSDSGQVLLSENIKYLPTWSTPV 394

RESULT 76
AAP94703
ID AAP94703 standard; protein; 524 AA.
XX
AC AAP94703;
XX
DT 25-MAR-2003 (revised)
DT 22-MAR-1991 (first entry)
XX
DE Sequence encoded by T4 lymphocyte cDNA obtained from PBL clone lambda-203
-4.
XX
XX HIV; soluble T4; immunotherapeutic; prophylactic; diagnostic; AIDS; ARC.
XX
XX Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 67 /note= "AA DESIGNATED NUMBER -23"
FT Misc-difference 90 /note= "AA DESIGNATED NUMBER 1"
FT Misc-difference 92 /note= "MATURE N-TERMINUS"
FT
XX
XX WC8901940-A.
XX
XX 09-MAR-1989.
XX
XX 01-SEP-1988; 88MO-US002940.
XX
XX 04-SEP-1987; 87US-00094322.
XX 07-JAN-1988; 88US-00141649.
XX
XX (BIOG ) BIOGEN INC.
XX
XX Fisher RA, Gilbert W, Sato VL, Flavell RA, Maganore JM;
XX
XX WPI; 1989-085519/11.
XX
XX N-PSDB; AAN90642.
XX
XX
XX DNA sequences coding for soluble T4-like polypeptide(s) - used in
XX immunotherapeutic and immunosuppressive compns. and for preventing,
XX treating or detecting AIDS.
XX
XX Disclosure; Fig 3; 207pp; English.
XX
XX The polypeptides encoded are useful in immunotherapeutic, prophylactic
XX and diagnostic compns. They can be used to purify HIV from a sample. The
XX soluble T4 protein-based compns. are useful in treating immunodeficient
XX patients suffering from diseases caused by agents whose primary targets
XX are T4+ lymphocytes. They can be used for preventing, treating or
XX detecting AIDS, ARC and HIV infection. (Updated on 25-MAR-2003 to correct
XX PR field.)
XX
XX Sequence 524 AA:
XX
XX Query Match 58.4%; Score 1993; DB 1; Length 524;
XX Best Local Similarity 99.0%; Pred. No. 1.2e-99;
XX Matches 390; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 MNRGVPFRHLLLVQLALLPAAATGKRVVLGKGGDTVELTCTASQKKSIQFHMKNNSQIK 60
DB 67 MNRGVPFRHLLLVQLALLPAAATGKRVVLGKGGDTVELTCTASQKKSIQFHMKNNSQIK 126

```

```

QY 61 ILGNQGSFLTKGSPSKLNDRAADRSRLMDQGNFLLIKNLKLTEDSDTYICEVDQKEEYQL 120
DB 127 ILGNQGSFLTKGSPSKLNDRAADRSRLMDQGNFLLIKNLKLTEDSDTYICEVDQKEEYQL 186
QY 121 LVFGLTANSDBTHLLQGOSLTTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLEIQQSG 180
DB 187 LVFGLTANSDBTHLLQGOSLTTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLEIQQSG 246
QY 181 TWTCTVLQNKQKVEERKIDIVLAFOKASSIYKKEGEQVEFSFPLATVTEKLTGSGELMW 240
DB 247 TWTCTVLQNKQKVEERKIDIVLAFOKASSIYKKEGEQVEFSFPLATVTEKLTGSGELMW 306
QY 241 QAEARASSKSWITTFDLKKNKVEVSKRVTQDPKLQMGKPLHLTLPOALPQVAGSGULTLA 300
DB 307 QAEARASSKSWITTFDLKKNKVEVSKRVTQDPKLQMGKPLHLTLPOALPQVAGSGULTLA 366
QY 301 LEAKTGKTLHQEVNLYVMRATOLQKNLTCEVWGPTSPKMLSLKLNKKAQVSKREKPVV 360
DB 367 LEAKTGKTLHQEVNLYVMRATOLQKNLTCEVWGPTSPKMLSLKLNKKAQVSKREKPVV 426
QY 361 LNPEAGMOCCLSDSGQVLLSENIKYLPTWSTPV 394
DB 427 LNPEAGMOCCLSDSGQVLLSENIKYLPTWSTPV 460

RESULT 77
AAR07640
ID AAR07640 standard; protein; 2458 AA.
XX
XX AAR07640;
XX
AC AAR07640;
XX
DT 31-OCT-2002 (revised)
DT 20-DEC-1990 (first entry)
XX
DE Deduced protein sequence of p170-2 comprising T4.
XX
XX plasmid p170-2; soluble T4 protein; AIDS; ARC; HIV.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
XX FT Protein 400..858
XX FT /label= T4 surface glycoprotein
XX
XX PN MO9008198-A.
XX
XX 26-JUL-1990.
XX
XX 18-JAN-1989; 89US-00300096.
XX
XX 18-JAN-1989; 89US-00300096.
XX
XX (HARD ) HARVARD COLLEGE.
XX
XX Letvin NA;
XX
XX WPI; 1990-254040/33.
XX N-PSDB; AAQ05607.
XX
XX Treating or preventing AIDS, ARC or HIV infection - by administering an
XX immunologically effective amt. of soluble T4 protein.
XX
XX Disclosure; Fig 1; 121pp; English.
XX
XX Entire sequence from T4-encoding plasmid p170-2. It is almost identical
XX to the sequence published by Maddon et al.(1985) with the exception of
XX three codon changes. At T4 amino acid residue 3, (posn.403 of entire
XX sequence) Lys is encoded in stead of Asn. At posn. 64, (posn.464) Arg
XX replaces Trp and at posn. 231, (posn. 631) Ser replaces Phe. Soluble T4
XX can be produced by truncating the CDS to remove the transmembrane and
XX cytoplasmic domains. The soluble forms may be modified to increase their
XX immunogenicity by addition of an adjuvant such as incomplete Freund's
XX adjuvant. The T4 interferes with HIV/T4 interaction and elicits anti-

```

soluble T4 antibody production. See also AA005608. (Updated on 31-OCT-2002 to add missing OS field.)

Sequence 2458 AA;

Query Match 58.4%; Score 1993; DB 2; Length 2458;
Best Local Similarity 99.0%; Pred. No. 6e-99;
Matches 390; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 MNRGVPFRHLVLTQALLPAATQGNKVVLGKGDVETLTCTASOKKSIQPHMKNNOIK 60
DB 400 MNRGVPFRHLVLTQALLPAATQGNKVVLGKGDVETLTCTASOKKSIQPHMKNNOIK 459
QY 61 ILGNQGSFLTKGPKLNDNRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEVOL 120
DB 460 ILGNQGSFLTKGPKLNDNRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEVOL 519
QY 121 LVFGLTANSPTHLTQGSGLTTLTSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB 520 LVFGLTANSPTHLTQGSGLTTLTSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 579
QY 181 TWCTVLOQOKKVEFKIDIVLAFOKASSIVYKKEGQVEFSFPLATFVEKLTGSGELMW 240
DB 580 TWCTVLOQOKKVEFKIDIVLAFOKASSIVYKKEGQVEFSFPLATFVEKLTGSGELMW 639
QY 241 QAERASSSKSWITFDLNKKEVSVKRYTODPKLOMGKKLPLHLTLPOLPOYAGSGNLTLA 300
DB 640 QAERASSSKSWITFDLNKKEVSVKRYTODPKLOMGKKLPLHLTLPOLPOYAGSGNLTLA 699
QY 301 LEAKTGKLGHEVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKLENKEAKVSKREKPVW 360
DB 700 LEAKTGKLGHEVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKLENKEAKVSKREKPVW 759
QY 361 LNPEAGMWQCLSDSGVLLSNIKVLPTWSTPV 394
DB 760 LNPEAGMWQCLSDSGVLLSNIKVLPTWSTPV 793

RESULT 78

AA04031
ID AAR04031 standard; protein; 2458 AA.

AC AAR04031;
XX
DT 25-MAR-2003 (revised)
DT 31-OCT-2002 (revised)
DT 29-MAY-1990 (first entry)
XX
DE Full length T4 encoded by plasmid p170-2.
XX
KM Soluble T4; p170-2; anti-retroviral agent; AIDS; ARC; HIV; AZT.
XX
OS Synthetic.
XX
XX Key Location/Qualifiers
FH MISC-difference 423
FT MISC-difference 425 /note= "TTP of Madden et al replaced by Arg"
FT MISC-difference 425 /note= "Asp of Madden et al replaced by Lys"
FT MISC-difference 653 /note= "Phe of Madden et al replaced by Ser"
XX
PN MO8911860-A.
XX
XX 14-DEC-1989.
XX
XX 08-JUN-1989; 89MO-US002453.
XX
XX 10-JUN-1988; 88US-00204645.
XX 20-APR-1989; 89US-00341080.
XX
XX (BIOJ) BIOGEN NV INC.
XX (GEHO) GEN HOSPITAL CORP.

PA (BIOJ) BIOGEN INC.
PA (BIOJ) BIOGEN INC.
XX
XX

PI Fisher RA, Schooley RT, Hirsch MS, Johnson VA, Walker BD;
XX

DR WPI; 1990-007302/01.
XX

DR N-PSDB; AA003005.
XX

PT Combinations of soluble T4 protein and anti-retroviral agent - having
PT synergistic activity in treatment and prevention of AIDS, arc and HIV
PT infection.
XX

PS Disclosure; Fig 1; 100pp; English.
XX

CC The sequence differs from that determined by PJ Madden et al., [Cell, 42
CC pp. 93-104 (1985)] in three places due to three nucleotide substitutions.
CC The Asp reported at position 3 by Madden et al. was the result of a
CC sequencing error (DR Litzman et al. Cell, 55, p.541 (1988)). X = stop
CC codon. The sequence was deduced from the cDNA insert of p170-2. Soluble
CC T4 constructs may be produced by truncating this sequence to give
CC fragments from position 400 to 799, removing the transmembrane and
CC intracytoplasmic domains whilst retaining the extracellular region
CC responsible for HIV binding. The sol. T4 is combined with an anti-viral
CC agent such as AZT. See also AA003006. (Updated on 31-OCT-2002 to add
CC missing OS field.) (Updated on 25-MAR-2003 to correct PA field.)
XX

SO Sequence 2458 AA;

Query Match 58.4%; Score 1993; DB 2; Length 2458;
Best Local Similarity 99.0%; Pred. No. 6e-99;
Matches 390; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 MNRGVPFRHLVLTQALLPAATQGNKVVLGKGDVETLTCTASOKKSIQPHMKNNOIK 60
DB 400 MNRGVPFRHLVLTQALLPAATQGNKVVLGKGDVETLTCTASOKKSIQPHMKNNOIK 459
QY 61 ILGNQGSFLTKGPKLNDNRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEVOL 120
DB 460 ILGNQGSFLTKGPKLNDNRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEVOL 519
QY 121 LVFGLTANSPTHLTQGSGLTTLTSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB 520 LVFGLTANSPTHLTQGSGLTTLTSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 579
QY 181 TWCTVLOQOKKVEFKIDIVLAFOKASSIVYKKEGQVEFSFPLATFVEKLTGSGELMW 240
DB 580 TWCTVLOQOKKVEFKIDIVLAFOKASSIVYKKEGQVEFSFPLATFVEKLTGSGELMW 639
QY 241 QAERASSSKSWITFDLNKKEVSVKRYTODPKLOMGKKLPLHLTLPOLPOYAGSGNLTLA 300
DB 640 QAERASSSKSWITFDLNKKEVSVKRYTODPKLOMGKKLPLHLTLPOLPOYAGSGNLTLA 699
QY 301 LEAKTGKLGHEVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKLENKEAKVSKREKPVW 360
DB 700 LEAKTGKLGHEVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKLENKEAKVSKREKPVW 759
QY 361 LNPEAGMWQCLSDSGVLLSNIKVLPTWSTPV 394
DB 760 LNPEAGMWQCLSDSGVLLSNIKVLPTWSTPV 793

RESULT 79

AA020151
ID AAR20151 standard; protein; 399 AA.

AC AAR20151;
XX
XX
DT 25-MAR-2003 (revised)
DT 31-MAR-1992 (first entry)
XX
XX Chimpanzee sol. CD4 encoded by pSC200.
XX
XX Human immunodeficiency virus; HIV; gp 120; AIDS; ARC; glycoprotein;

KM acquired immune deficiency syndrome; AIDS related complex;
 KW T helper lymphocytes.
 XX Pan troglodytes.
 OS
 FH Key Location/Qualifiers
 FT Peptide 1..25
 FT /label= signal_sequence
 XX
 XX MO9118618-A.
 XX
 PD 12-DEC-1991.
 XX
 PF 25-MAY-1990; 90US-00529186.
 XX
 PR 25-MAY-1990; 90US-00529186.
 XX
 PA (BIOU) BIOGEN INC.
 XX
 PI Fisher RA, Heseion C, Burkly LC;
 XX
 DR WPI; 1992-007200/01.
 DR N-PSDB; AAQ20326.
 XX
 PT New immuno-therapeutic human CD4 variants and derive. - elicit AB
 PT production to HIV gp.120, useful in treating, preventing and diagnosing
 PT AIDS, ARC and HIV infections.
 XX
 PS Claim 15; Fig 21; 179pp; English.
 XX
 CC The sequence was deduced from the DNA sequence of clone pSQ200 which was
 CC obt'd. by cloning using a reverse transcriptase/PCR amplification
 CC procedure with a non-specific oligo dt to prime first strand synthesis.
 CC The PCR fragments were ligated into sequencing plasmid pNN08 (see
 CC AAQ20326) The DNA sequence can be used to express recombinant soluble CD4
 CC and analogues for use in diagnosis and treatment of diseases caused by
 CC infective agents whose primary targets are T4+ lymphocytes. See also
 CC AAR20148-R20155 and AAR21078. (Updated on 25-MAR-2003 to correct PA
 CC field.)
 CC
 SQ Sequence 399 AA;

Query Match 58.3%; Score 1991; DB 2; Length 399;
 Best Local Similarity 98.5%; Pred. No. 1.2e-99;
 Matches 388; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 MNRGVPRHLLVLTALPAATQGNKVVLGKKGDTVELTCTASQKKSIOFHKNSNQIK 60
 DB 1 MNRGVPRHLLVLTALPAATQGNKVVLGKKGDTVELTCTASQKKSIOFHKNSNQIK 60
 QY 61 ILGNQGSFLTKGPSKLNDRADRSRLMDQGNFLLIKNLKIEDSDTYICEVGDQKEVOL 120
 DB 61 ILGNQGSFLTKGPSKLNDRADRSRLMDQGNFLLIKNLKIEDSDTYICEVGDQKEVOL 120
 QY 121 LVFGLTANSPTHLLOGSITLTLESPPGSSPSVQCSPRGKNIQGGKTLVSQLELDQSG 180
 DB 121 LVFGLTANSPTHLLOGSITLTLESPPGSSPSVQCSPRGKNIQGGKTLVSQLELDQSG 180
 QY 121 LVFGLTANSPTHLLOGSITLTLESPPGSSPSVQCSPRGKNIQGGKTLVSQLELDQSG 180
 DB 121 LVFGLTANSPTHLLOGSITLTLESPPGSSPSVQCSPRGKNIQGGKTLVSQLELDQSG 180
 QY 181 TWCTVLYONQKVEFKIDIVLAFOKASSIVYKKEGEQVEFSFPLAFVTEKLTGSGELMW 240
 DB 181 TWCTVLYONQKVEFKIDIVLAFOKASSIVYKKEGEQVEFSFPLAFVTEKLTGSGELMW 240
 QY 241 QAEBASSSKSWITFDLKNKEVSVKRVTPDKLQMGKCLPLHLTPQALPOYAGSGNLTLA 300
 DB 241 QAEBASSSKSWITFDLKNKEVSVKRVTPDKLQMGKCLPLHLTPQALPOYAGSGNLTLA 300
 QY 301 LEAKTGLHOEVNLVYVNRATQLOKNLTCEVWGPSPKLMSTLKENKEAVSRREKRVWV 360
 DB 301 LEAKTGLHOEVNLVYVNRATQLOKNLTCEVWGPSPKLMSTLKENKEAVSRREKRVWV 360
 QY 361 LNPBAGMWOCILSDSGOVLLESNIKVLPTWSTPV 394
 DB 361 LNPBAGMWOCILSDSGOVLLESNIKVLPTWSTPV 394

RESULT 80
 AAR11285
 ID AAR11285 standard; protein; 458 AA.
 XX
 XX AAR11285;
 AC
 XX 25-MAR-2003 (revised)
 DT 09-JAN-2003 (revised)
 DT 29-APR-1991 (first entry)
 XX
 DE gp120 binding protein.
 XX
 KW Human; CD4; AIDS; HIV1; SIV; gp120.
 XX
 OS Unidentified.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 59
 FT /label= Thr or Ile
 FT Misc-difference 93
 FT /label= Thr or Pro
 XX
 PN EP414178-A.
 XX
 PD 27-FEB-1991.
 XX
 PF 18-AUG-1990; 90EP-00115877.
 XX
 PR 23-AUG-1989; 89US-0039782.
 XX
 PA (GENO) GEN HOSPITAL CORP.
 XX
 PI Seed B, Camerini D;
 XX
 DR WPI; 1991-059419/09.
 DR N-PSDB; AAQ10887.
 XX
 XX New non-human primate and human CD4 or gp120 molecules - used to treat
 PT HIV or SIV and immunoglobulin and gp120 binding molecules from new fusion
 PT proteins.
 XX
 PS Claim 17; Page 57; 87pp; English.
 XX
 CC The fragment from residues 1-134 is also independently claimed. The sub-
 CC fragment (and the complete polypeptide) can bind to HIV gp120. See also
 CC AAQ10885-6, AAQ10888. (Updated on 09-JAN-2003 to add missing OS field.)
 CC (Updated on 25-MAR-2003 to correct PA field.)
 XX
 SQ Sequence 458 AA;

Query Match 58.3%; Score 1991; DB 2; Length 458;
 Best Local Similarity 98.5%; Pred. No. 1.3e-99;
 Matches 388; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 MNRGVPRHLLVLTALPAATQGNKVVLGKKGDTVELTCTASQKKSIOFHKNSNQIK 60
 DB 1 MNRGVPRHLLVLTALPAATQGNKVVLGKKGDTVELTCTASQKKSIOFHKNSNQIK 60
 QY 61 ILGNQGSFLTKGPSKLNDRADRSRLMDQGNFLLIKNLKIEDSDTYICEVGDQKEVOL 120
 DB 61 ILGNQGSFLTKGPSKLNDRADRSRLMDQGNFLLIKNLKIEDSDTYICEVGDQKEVOL 120
 QY 121 LVFGLTANSPTHLLOGSITLTLESPPGSSPSVQCSPRGKNIQGGKTLVSQLELDQSG 180
 DB 121 LVFGLTANSPTHLLOGSITLTLESPPGSSPSVQCSPRGKNIQGGKTLVSQLELDQSG 180
 QY 121 LVFGLTANSPTHLLOGSITLTLESPPGSSPSVQCSPRGKNIQGGKTLVSQLELDQSG 180
 DB 121 LVFGLTANSPTHLLOGSITLTLESPPGSSPSVQCSPRGKNIQGGKTLVSQLELDQSG 180
 QY 181 TWCTVLYONQKVEFKIDIVLAFOKASSIVYKKEGEQVEFSFPLAFVTEKLTGSGELMW 240
 DB 181 TWCTVLYONQKVEFKIDIVLAFOKASSIVYKKEGEQVEFSFPLAFVTEKLTGSGELMW 240
 QY 241 QAEBASSSKSWITFDLKNKEVSVKRVTPDKLQMGKCLPLHLTPQALPOYAGSGNLTLA 300

```

Db      241 EAERASSSKSWITFDLKNKEVSVKRVTDPKLQMGKLP.LHLTL.PQALPYAGSGNLTLA 300
Qy      301 LEAKTGKHOEVLVWVRATOLQKLTCEVWGPTSPKLMSTLKENKAKVSKREKPYWV 360
Db      301 LEAKTGKHOEVLVWVRATOLQKLTCEVWGPTSPKLMSTLKENKAKVSKREKAVWV 360
Qy      361 LNPEAGMWQCLSDSGVLLSNIKVLPTWSTPV 394
Db      361 LNPEAGMWQCLSDSGVLLSNIKVLPTWSTPV 394

```

RESULT 81

```

AAR10988 ID AAR10988 standard; protein; 458 AA.
XX
AC AAR10988;
XX
DT 25-MAR-2003 (revised)
DT 29-APR-1991 (first entry)
XX
DE Chimpanzee CD4 protein.
XX
KM Chimpanzee; CD4; AIDS; HIV1; SIV.
XX
OS Pan troglodytes.
XX
FH Key Location/Qualifiers
FT Protein 26
FT /label= mature CD4
XX
PN EP414178-A.
XX
PD 27-FEB-1991.
XX
PF 18-AUG-1990; 90EP-00115877.
XX
PR 23-AUG-1989; 89US-00397782.
XX
PA (GENO ) GEN HOSPITAL CORP.
XX
PI Seed B, Camerini D;
XX
DR WPI; 1991-059419/09.
DR N-PSDB; AAQ10886.
XX
PT New non-human primate and human CD4 or gp120 molecules - used to treat
PT HIV or SIV and immunoglobulin and gp120 binding molecules from new fusion
PT proteins.
XX
PS Claim 4; Page 45; 87pp; English.
XX
CC The CD4 protein or HIV gp120-binding fragments of it are used to detect
CC and treat HIV and SIV infection. Animals which can be treated include
CC humans, baboons, orang-utans, chimpanzees, gorillas and rhesus monkeys.
CC The chimpanzee CD4 is 99 per cent homologous to its human counterpart,
CC possessing 5 amino acid substitutions in the 433 residue predicted mature
CC polypeptide. See also AAQ10885, AAQ10887-8. (Updated on 25-MAR-2003 to
CC correct PA field.)
XX
SQ Sequence 458 AA;

```

Query Match 58.3%; Score 1991; DB 2; Length 458;

Beet Local Similarity 98.5%; Pred. No. 1.3e-99;

Matches 388; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

Qy      1 MNRGVFPHLLVLTQALLPATQGNKVVLGKGGTVELTCTASOKSIOFHWKNSQIK 60
Db      1 MNRGVFPHLLVLTQALLPATQGNKVVLGKGGTVELTCTASOKSIOFHWKNSQIK 60
Qy      61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNP.LIKVLKTIKEDSDTYICEVEDQKEEVOL 120
Db      61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNP.LIKVLKTIKEDSDTYICEVEDQKEEVOL 120

```

```

Qy      121 LVFGLTANSPDTHLLQGSLLTTLSPGSSPSVQCRRPRGKNIQGGKTLSSVSOLELQDSG 180
Db      121 LVFGLTANSPDTHLLQGSLLTTLSPGSSPSVQCRRPRGKNIQGGKTLSSVSOLELQDSG 180
Qy      181 TWFTCTVLQNKQKVEFKIDIVLAFQKASSIVYKKEGQVEFSPEPLAFTEKLTGSGELMW 240
Db      181 TWFTCTVLQNKQKVEFKIDIVLAFQKASSIVYKKEGQVEFSPEPLAFTEKLTGSGELMW 240
Qy      241 QAERASSSKSWITFDLKNKEVSVKRVTDPKLQMGKLP.LHLTL.PQALPYAGSGNLTLA 300
Db      241 QAERASSSKSWITFDLKNKEVSVKRVTDPKLQMGKLP.LHLTL.PQALPYAGSGNLTLA 300
Qy      301 LEAKTGKHOEVLVWVRATOLQKLTCEVWGPTSPKLMSTLKENKAKVSKREKPYWV 360
Db      301 LEAKTGKHOEVLVWVRATOLQKLTCEVWGPTSPKLMSTLKENKAKVSKREKAVWV 360
Qy      361 LNPEAGMWQCLSDSGVLLSNIKVLPTWSTPV 394
Db      361 LNPEAGMWQCLSDSGVLLSNIKVLPTWSTPV 394

```

RESULT 82

```

AAR20150 ID AAR20150 standard; protein; 400 AA.
XX
AC AAR20150;
XX
DT 25-MAR-2003 (revised)
DT 31-MAR-1992 (first entry)
XX
DE Chimpanzee sol. CD4 sequence from psQ205.
XX
KM Human immunodeficiency virus; HIV; gp 120; AIDS; ARC; glycoprotein;
KM acquired immune deficiency syndrome; AIDS related complex;
KM T helper lymphocytes.
XX
OS Pan troglodytes.
XX
FH Key Location/Qualifiers
FT Peptide 1..25
FT /label= signal_sequence
XX
PN W09118618-A.
XX
PD 12-DEC-1991.
XX
PF 25-MAY-1990; 90US-00529186.
XX
PR 25-MAY-1990; 90US-00529186.
XX
PA (BIOU ) BIOGEN INC.
XX
PI Fisher RA, Hession C, Burkly LC;
XX
DR WPI; 1992-007200/01.
DR N-PSDB; AAQ20325.
XX
PT New immuno-therapeutic human CD4 variants and derive. - elicit AB
PT production to HIV gp.120, useful in treating, preventing and diagnosing
PT AIDS, ARC and HIV infections.
XX
PS Claim 15; Fig 20; 179pp; English.
XX
CC The sequence was deduced from the DNA sequence of clone psQ205 which was
CC obd. by cloning using a reverse transcriptase/PCR amplification
CC procedure. The DNA sequence can be used to express recombinant soluble
CC CD4 and analogues for use in diagnosis and treatment of diseases caused
CC by infective agents whose primary targets are T4+ lymphocytes. See also
CC AAR20148-R20155 and AAR21078. (Updated on 25-MAR-2003 to correct PA
CC field.)
XX
SQ Sequence 400 AA;

```

Query Match 58.2%; Score 1986; DB 2; Length 400;
 Best Local Similarity 98.2%; Pred. No. 2.2e-99;
 Matches 387; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 MNRGVFPRHLLVLTALPAATQGNKVLGKKGDTVELCTASQKKSIOFHMKNNOIK 60
 |||||||
 DB 1 MNRGVFPRHLLVLTALPAATQGNKVLGKKGDTVELCTASQKKSIOFHMKNNOIK 60
 |||||||

QY 61 ILGNQGSFLTKGPKSLNDRADSRSLMDQGNPPLIIKNLIKIDSDTYICEVEDQKEEYOL 120
 |||||||
 DB 61 ILGNQGSFLTKGPKSLNDRADSRSLMDQGNPPLIIKNLIKIDSDTYICEVEDQKEEYOL 120
 |||||||

QY 121 LVFGLTANSPTHTLLQGSLLTLTLESPGSSPVQCSPPKKNIOGKTLVSQLELDSDG 180
 |||||||
 DB 121 LVFGLTANSPTHTLLQGSLLTLTLESPGSSPVQCSPPKKNIOGKTLVSQLELDSDG 180
 |||||||

QY 181 TWICTVLQONQKVEFKIDIVLAFQKASSIVYKKEGEQVEFSPPLAFYVEKLTGSGELMW 240
 |||||||
 DB 181 TWICTVLQONQKVEFKIDIVLAFQKASSIVYKKEGEQVEFSPPLAFYVEKLTGSGELMW 240
 |||||||

QY 241 QAERASSSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPLHLTLPQALPOYAGSGNLTLA 300
 |||||||
 DB 241 QAERASSSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPLHLTLPQALPOYAGSGNLTLA 300
 |||||||

QY 301 LEAKTGKLEHVENLVVMRATQLOKNTCEVWGPTSPKMLSLKLENKEAKVSKREKPVWV 360
 |||||||
 DB 301 LEAKTGKLEHVENLVVMRATQLOKNTCEVWGPTSPKMLSLKLENKEAKVSKREKPVWV 360
 |||||||

QY 361 LNPEAGMOCCLSDSGOVLLESNIKVLPTWSTPV 394
 |||||||
 DB 361 LNPEAGMOCCLSDSGOVLLESNIKVLPTWSTPV 394
 |||||||

RESULT 83
 AAR04910 standard; protein; 458 AA.

XX AAR04910;
 XX 31-OCT-2002 (revised)
 DT 02-OCT-1990 (first entry)
 XX
 DE T4 protein as encoded by p170.2.
 XX
 KM T4 protein; immunotoxin; Pseudomonas endotoxin A; AIDS; HIV; ARC;
 KM angiogenin; fusion protein.
 XX
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..26
 FT Protein /label= signal peptide
 FT Protein 27..458
 FT /label= T4 protein
 FT Misc-difference replace (87 W)
 FT /note= "differs from Madden et al"
 FT Misc-difference replace (254 F)
 FT /note= "differs from Madden et al"
 XX
 PN WO9004414-A.
 XX
 PD 03-MAY-1990.
 XX
 PF 18-OCT-1988; 88US-00259355.
 XX
 PR 18-OCT-1988; 88US-00259355.
 XX
 PA (BIO) BIOGEN INC.
 XX
 PI Meade HM, Lobb RR, Gates LL, Winkler G;
 XX
 DR WPI, 1990-163876/21.

DR N-PSDB; AAR04555.
 XX
 XX New immunotoxin contg. soluble T4 protein components and toxin - esp.
 PT Pseudomonas endotoxin A, for treating or controlling AIDS and related
 PT conditions, and new DNA sequences.
 PS
 XX Disclosure; Page 7; -pp; English.

CC The T4 protein encoded by p170.2 is almost identical to that reported by
 CC P.J. Madden et al [Cell, 42, pp 93-104 (1985)]. The Madden sequence was
 CC revised in 1988 to correct a DNA sequencing error at AA 3 (corrected from
 CC Asp to Lys; see M12807 in Genbank). The DNA may be truncated (to remove
 CC transmembrane and intercellular regions) and/or modified by SDM, pref. so
 CC the prod. extends from AAs 3-183 of the mature protein. This DNA can then
 CC be ligated to a toxin DNA esp. angiogenin, or a fragment of Pseudomonas
 CC exotoxin A contg. the translocation and ADP-ribosylation domains. The
 CC hybrid DNA can then be inserted into an expression vector and used to
 CC produce recombinant fusion protein which is useful for preventing or
 CC treating AIDS, ARC, and HIV infections. The T4 protein is an HIV receptor
 CC which binds to virus or to infected cells carrying the gp120/160 marker
 CC antigen, so provides v. specific targeting with minimal damage to non-
 CC target cells. (Updated on 31-OCT-2002 to add missing OS field.)
 CC
 XX
 SQ Sequence 458 AA;

Query Match 58.1%; Score 1985; DB 2; Length 458;
 Best Local Similarity 98.5%; Pred. No. 2.8e-99;
 Matches 387; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 MNRGVFPRHLLVLTALPAATQGNKVLGKKGDTVELCTASQKKSIOFHMKNNOIK 60
 |||||||
 DB 1 MNRGVFPRHLLVLTALPAATQGNKVLGKKGDTVELCTASQKKSIOFHMKNNOIK 60
 |||||||

QY 61 ILGNQGSFLTKGPKSLNDRADSRSLMDQGNPPLIIKNLIKIDSDTYICEVEDQKEEYOL 120
 |||||||
 DB 61 ILGNQGSFLTKGPKSLNDRADSRSLMDQGNPPLIIKNLIKIDSDTYICEVEDQKEEYOL 120
 |||||||

QY 121 LVFGLTANSPTHTLLQGSLLTLTLESPGSSPVQCSPPKKNIOGKTLVSQLELDSDG 180
 |||||||
 DB 121 LVFGLTANSPTHTLLQGSLLTLTLESPGSSPVQCSPPKKNIOGKTLVSQLELDSDG 180
 |||||||

QY 181 TWICTVLQONQKVEFKIDIVLAFQKASSIVYKKEGEQVEFSPPLAFYVEKLTGSGELMW 240
 |||||||
 DB 181 TWICTVLQONQKVEFKIDIVLAFQKASSIVYKKEGEQVEFSPPLAFYVEKLTGSGELMW 240
 |||||||

QY 241 QAERASSSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPLHLTLPQALPOYAGSGNLTLA 300
 |||||||
 DB 241 QAERASSSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPLHLTLPQALPOYAGSGNLTLA 300
 |||||||

QY 301 LEAKTGKLEHVENLVVMRATQLOKNTCEVWGPTSPKMLSLKLENKEAKVSKREKPVWV 360
 |||||||
 DB 301 LEAKTGKLEHVENLVVMRATQLOKNTCEVWGPTSPKMLSLKLENKEAKVSKREKPVWV 360
 |||||||

QY 361 LNPEAGMOCCLSDSGOVLLESNIKVLPTWSTPV 394
 |||||||
 DB 361 LNPEAGMOCCLSDSGOVLLESNIKVLPTWSTPV 394
 |||||||

RESULT 84
 AAP93010 standard; protein; 399 AA.

XX AAP93010;
 XX 25-MAR-2003 (revised)
 DT 03-AUG-1992 (first entry)
 XX
 DE Genetic construct which encodes CD4 linked to human IgM at the Met2 site
 DE upstream of the CH1 region (fusion protein CD4Hnu).
 XX
 KM Fusion protein; immunoglobulin-like molecule; HIV; SIV; therapy;
 KM diagnosis; CD4; gp120; binding fragment; glycoprotein; variable region.
 XX

OS Homo sapiens.
 XX EP325262-A.
 XX 26-JUL-1989.
 XX 20-JAN-1989; 89EP-00100913.
 XX 22-JAN-1988; 88US-00147351.
 XX (GENO) GEN HOSPITAL CORP.
 XX Seed B;
 XX WPI; 1989-214472/30.
 XX N-PSDB; AAN90358.
 DR Immunoglobulin-CD4 fusion proteins - used for treating HIV or SIV
 FT Infection or detecting HIV or SIV in sample.
 PS Example; Table 3, Page 34-40; 68pp; English.
 XX The fusion protein genes of the invention pref. comprises cDNA sequences
 CC which encode CD4 or a fragment which binds gp120 ligated to an expression
 CC plasmid which encodes an antibody in which the variable region of the
 CC gene has been deleted (see WO87-02671). The CD4 portion of the fusion
 CC protein may comprise the complete CD4 sequence, the 370 AA extracellular
 CC region and the membrane spanning domain, or the extracellular region. The
 CC Ig heavy chain is pref. from 19M, 19G1 or 19G3. The following are
 CC specifically claimed: fusion proteins CD4H1ambdA, CD4Mmu, CD4Pmu,
 CC CD4E1ambdA, and CD4Mmu (No. 67608), CD4P1ambdA (No. 67609) and
 CC PCD4E1ambdA (No. 67610). (Updated on 25-MAR-2003 to correct PA field.)
 XX Sequence 399 AA;
 SQ

Query Match 58.1%; Score 1982.5; DB 1; Length 399;
 Best Local Similarity 94.9%; Pred. No. 3.4e-99;
 Matches 391; Conservative 1; Mismatches 7; Indels 13; Gaps 1;

QY 1 MNRGVPFRHLLVQLALLPAATQGNKVLGKGDVVELTCTASQKSIQFHMKNNOIK 60
 DB 1 MNRGVPFRHLLVQLALLPAATQGNKVLGKGDVVELTCTASQKSIQFHMKNNOIK 60
 QY 61 ILNGSGFLTKGPKSLNDRADSRSLMDQGNPILITKLTIEDSDTYICEVEDQKEVQL 120
 DB 61 ILNGSGFLTKGPKSLNDRADSRSLMDQGNPILITKLTIEDSDTYICEVEDQKEVQL 120
 QY 121 LVFGLTANSPTHLLOGSLTLTLSPGSSPSVOCRSRPRGNIOGKTLVSQLELDDSG 180
 DB 121 LVFGLTANSPTHLLOGSLTLTLSPGSSPSVOCRSRPRGNIOGKTLVSQLELDDSG 180
 QY 121 LVFGLTANSPTHLLOGSLTLTLSPGSSPSVOCRSRPRGNIOGKTLVSQLELDDSG 180
 DB 121 LVFGLTANSPTHLLOGSLTLTLSPGSSPSVOCRSRPRGNIOGKTLVSQLELDDSG 180
 QY 181 TWCTCTVLONQKVEFKIDIVLAFQKASSIVYKKEGEQVEFSPLATFVEKLTGSGELMW 240
 DB 181 TWCTCTVLONQKVEFKIDIVLAFQKASSIVYKKEGEQVEFSPLATFVEKLTGSGELMW 240
 QY 241 QAERASSSKSWITPDLKNKEVSVKRVTDPLQNGKTLPLTLTPOALPOYAGSGNLTLA 300
 DB 241 QAERASSSKSWITPDLKNKEVSVKRVTDPLQNGKTLPLTLTPOALPOYAGSGNLTLA 300
 QY 301 LEATGKTLHQEVNLVVMKATOLQKLTCEVWGPTSPKMLSLKLENKEAKVSKKEKPYWV 360
 DB 301 LEATGKTLHQEVNLVVMKATOLQKLTCEVWGPTSPKMLSLKLENKEAKVSKKEKPYWV 360
 QY 361 LNPEAGMOCILSDSGQVLLSNIKVLPTWSTPVPCPAPEPKSCDKHTTCE 412
 DB 361 LNPEAGMOCILSDSGQVLLSNIKVLPTWSTPV-----HADPE 399

RESULT 85
 ID AAR20154
 AC AAR20154; standard; protein; 400 AA.
 AC AAR20154;

XX 24-OCT-2003 (revised)
 DT 25-MAR-2003 (revised)
 DT 31-MAR-1992 (first entry)
 XX Sol. rheus-human chimeric CD4 encoded by pDG100.
 DE Human immunodeficiency virus; HIV; gp 120; AIDS; ARC; glycoprotein;
 KW acquired immune deficiency syndrome; AIDS related complex; monkey;
 KW T helper lymphocytes.
 XX Macaca mulatta; (Rhesus).
 OS Homo sapiens.
 OS Chimeric.
 XX Key
 FT Region 1. .131
 FT Location/Qualifiers
 FT /note= "rheus CD4 encoded by pSQ146 (AAQ20328)"
 FT Peptide 1. .25
 FT /label= signal_sequence
 FT Region 132. .400
 FT /note= "human CD4 encoded by PBG391 (US8802940)"

XX WO9118618-A.
 XX 12-DEC-1991.
 XX 25-MAY-1990; 90US-00529186.
 XX 25-MAY-1990; 90US-00529186.
 XX (BIOJ) BIOGEN INC.
 XX Fisher RA, Hession C, Burklj LC;
 XX WPI; 1992-007200/01.
 DR N-PSDB; AAQ20329.
 XX New immuno-therapeutic human CD4 variants and derivs. - elicit AB
 FT production to HIV gp.120, useful in treating, preventing and diagnosing
 FT AIDS, ARC and HIV infections.

PS Claim 14; Fig 13; 179pp; English.
 CC The sequence is encoded by pDG100, chimeric plasmid prepd. from human and
 CC rheus derived CD4 DNA. The plasmid can be used to express recombinant
 CC sol. chimeric CD4 for use in diagnosis and treatment of diseases caused
 CC by infective agents whose primary targets are T4+ lymphocytes. See also
 CC AAR20148-R20155 and AAR21078. (Updated on 25-MAR-2003 to correct PA
 CC field.) (Updated on 24-OCT-2003 to standardise OS field)

SQ Sequence 400 AA;

Query Match 57.0%; Score 1946; DB 2; Length 400;
 Best Local Similarity 95.7%; Pred. No. 3.1e-97;
 Matches 377; Conservative 9; Mismatches 8; Indels 0; Gaps 0;

QY 1 MNRGVPFRHLLVQLALLPAATQGNKVLGKGDVVELTCTASQKSIQFHMKNNOIK 60
 DB 1 MNRGVPFRHLLVQLALLPAATQGNKVLGKGDVVELTCTASQKSIQFHMKNNOIK 60
 QY 61 ILNGSGFLTKGPKSLNDRADSRSLMDQGNPILITKLTIEDSDTYICEVEDQKEVQL 120
 DB 61 ILNGSGFLTKGPKSLNDRADSRSLMDQGNPILITKLTIEDSDTYICEVEDQKEVQL 120
 QY 121 LVFGLTANSPTHLLOGSLTLTLSPGSSPSVOCRSRPRGNIOGKTLVSQLELDDSG 180
 DB 121 LVFGLTANSPTHLLOGSLTLTLSPGSSPSVOCRSRPRGNIOGKTLVSQLELDDSG 180
 QY 181 TWCTCTVLONQKVEFKIDIVLAFQKASSIVYKKEGEQVEFSPLATFVEKLTGSGELMW 240
 DB 181 TWCTCTVLONQKVEFKIDIVLAFQKASSIVYKKEGEQVEFSPLATFVEKLTGSGELMW 240

QY 241 QAERASSSSKSNITFDLKNKEVSVKRVYTOPDKLQMGKKLPPLHLTPQALPOYAGSGNLTIA 300
 DB 241 QAERASSSSKSNITFDLKNKEVSVKRVYTOPDKLQMGKKLPPLHLTPQALPOYAGSGNLTIA 300
 QY 301 LEAKTGLHGVNLVVMRAATQLOKNLTCEVWGPTSPKMLSLKLENKAQVSKREKRVVW 360
 DB 301 LEAKTGLHGVNLVVMRAATQLOKNLTCEVWGPTSPKMLSLKLENKAQVSKREKRVVW 360
 QY 361 LNPBAGMOCCLSDSGVLLSNIKVLPWTSTPV 394
 DB 361 LNPBAGMOCCLSDSGVLLSNIKVLPWTSTPV 394

RESULT 86
 AAR1042
 ID AAR41042 standard; protein; 729 AA.
 XX
 AC AAR41042;
 XX
 DT 24-OCT-2003 (revised)
 DT 25-MAR-2003 (revised)
 DT 22-MAR-1994 (first entry)
 XX
 DE CD4-GBPH fusion protein.
 XX
 KM Merozoite; Glycophorin Binding Protein homologue; malaria; HIV; env;
 KM human immunodeficiency virus; envelope glycoprotein; hybrid protein;
 KM red blood cell; erythrocyte; AIDS.
 XX
 OS Homo sapiens.
 OS Plasmodium falciparum.
 OS Chimeric.
 XX
 FT Key Location/Qualifiers
 FT Region 1..371
 FT /note="residues 1-371 of CD4"
 FT Region 372..729
 FT /note="residues 70-427 of GBPH"
 FT
 PN WO9318160-A1.
 PD 16-SEP-1993.
 PF 10-MAR-1993; 93MO-GB000505.
 XX
 PR 11-MAR-1992; 92GB-00005276.
 PR 08-JUL-1992; 92GB-00014481.
 PR 24-JUL-1992; 92GB-00015829.
 PR 16-SEP-1992; 92GB-00019562.
 PR 03-MAR-1993; 93GB-00004311.
 XX
 PA (PREN/) PRENDERGAST K F.
 PI Prendergast KF;
 XX
 DR WPI; 1993-303474/38.
 XX
 PT Anti-viral fusion peptide(s) - comprise viral-binding component and
 PT malaria merozoite red cell binding component, for treating e.g. HIV, and
 PT hepatitis.
 XX
 PS Claim 7; Page 40-41; 69pp; English.
 XX
 CC The hybrid protein NH2-CD4(1-371)-GBPH(70-427)-COOH is a specifically
 CC claimed example of a fusion protein of the invention; it comprises at
 CC least part of the CD4 molecule fused to a peptide from a malarial
 CC parasite merozoite protein with affinity for red blood cells. The fusion
 CC protein can bind free HIV in the blood to red blood cells and
 CC consequently reduce viral titre, prevent transmission of the virus and
 CC improve safety of blood transfusions. (Updated on 25-MAR-2003 to correct
 CC PN field.) (Updated on 24-OCT-2003 to standardise OS field)

Query Match 56.5%; Score 1929.5; DB 2; Length 729;
 Best Local Similarity 63.9%; Pred. No. 4, 5e-96;
 Matches 434; Conservative 34; Mismatches 104; Indels 107; Gaps 18;

QY 24 OGNKVVVGKGGDTVELTCTASQKKSIOFPMKNSNOIKILGNQSPFLTKGPSKLNDRADR 83
 DB 1 OGNKVVVGKGGDTVELTCTASQKKSIOFPMKNSNOIKILGNQSPFLTKGPSKLNDRADR 60
 QY 84 RSLMDQGNFPLIIKNLIKEDSDTYICEVDQKEVQLVFGLTANSDTHLQGSITLTL 143
 DB 61 RSLMDQGNFPLIIKNLIKEDSDTYICEVDQKEVQLVFGLTANSDTHLQGSITLTL 120
 QY 144 ESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELQDSGTCTVLQNKQKVEKIDIVLVA 203
 DB 121 ESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELQDSGTCTVLQNKQKVEKIDIVLVA 180
 QY 204 FOKASSIYKKEGEOVEFSPLAFTVEKLTGSGELMWAERASSSKSNITFDLKNKEVSV 263
 DB 181 FOKASSIYKKEGEOVEFSPLAFTVEKLTGSGELMWAERASSSKSNITFDLKNKEVSV 240
 QY 264 KRVTDOPKLOMGKKLPPLHLTPQALPOYAGSGNLTIALEAKTGKLGHOEVNLVVMRAATQLO 323
 DB 241 KRVTDOPKLOMGKKLPPLHLTPQALPOYAGSGNLTIALEAKTGKLGHOEVNLVVMRAATQLO 300
 QY 324 KNLTCGVWGPTSPKMLSLKLENKAQVSKREKRVVNLNPEAGMOCCLSDSGVLLSN 383
 DB 301 KNLTCGVWGPTSPKMLSLKLENKAQVSKREKRVVNLNPEAGMOCCLSDSGVLLSN 360
 QY 384 IKVLPWTSTPV-----PCPAPRPSCKDHTCPPLGLGPSVFLPFPKPKDTL 430
 DB 361 IKVLPWTSTPVQYQYQADYSFRRESRVLAEGSTSKAKATL-----RRTKQTT 410
 QY 431 MISRPEVTGVVDVSHEDPEVKFMVYDGEVHNAKTKPREQYNST----- 478
 DB 411 LRSADPEQ--IMKWAADPEYRKHLNVLQILNN--TDPNDELTSADPEQIMKAYAA 466
 QY 479 ---YRVSVTLVLDHQMUNGKEYKCKVSNKALPAPLEKTSIKAKQPREPQYTLTPPSAD 535
 DB 467 DEPR--KHLNVLQY--ILNNTDNDVESSADP--EQQIMKA-----YAADP-- 508
 QY 536 ELTKQVSLTCLVKGFPSPDIAVWESNQOPEN-----NYKTPPVL-----DSD 580
 DB 509 EYRKHNVLVYQILNNTDND---ELETSADPEQIMKAYAADPEYRKHNVLVYQILNNTD 565
 QY 581 GSEFLYSKLTVDKSRMONGVFS-----SWHEALHNHYTKSLSLSPQLQD 629
 DB 566 SS-----EVEITSADPEQIMKAYAADPEYRKHNVLVYQIL--NHTDSSEV----- 608
 QY 630 ETCANQDELQGLWTTDP 648
 DB 609 ETSAD-PEQIMKAYAADP 626

RESULT 87
 AAR15149
 ID AAR15149 standard; protein; 458 AA.
 XX
 AC AAR15149;
 XX
 DT 25-MAR-2003 (revised)
 DT 24-FEB-1992 (first entry)
 XX
 DE CD4 coordinate system.
 DE Gelsolin; fusion protein; diagnosis; AIDS.
 XX
 OS Homo sapiens.
 XX
 FT Key Location/Qualifiers
 FT Domain 1..25
 FT /label= hydrophobic/secretory_signal
 FT Domain 26..132

Query	Match	Score	1921	DB 2	Length	458
Query	Local Similarity	95.2%	Pred. No. 8e-96			
Matches	375	Conservative	7	Mismatches	12	Indels
						Gaps
						0
1	MNRGPPFHLILVQLALPPAATGQNKVLLGKGGDTVLTCTAGQKSIQFHWKNSNQIK	60				
2	ILGNQSGFLTGPCSKLNRADSRSLMDQGNFPLTIKMLKIEDSTYICEVEDQKEEYQL	120				
3	ILGNQSGFLTGPCSKLNRADSRSLMDQGNFPLTIKMLKIEDSTYICEVEDQKEEYQL	120				
4	LVFGITANSDBTHLLOGQSLTTLTLESPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG	180				
5	LVFGITANSDBTHLLOGQSLTTLTLESPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG	180				
6	LVFGITANSDBTHLLOGQSLTTLTLESPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG	180				
7	LVFGITANSDBTHLLOGQSLTTLTLESPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG	180				
8	LVFGITANSDBTHLLOGQSLTTLTLESPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG	180				
9	LVFGITANSDBTHLLOGQSLTTLTLESPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG	180				
10	LVFGITANSDBTHLLOGQSLTTLTLESPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG	180				
11	LVFGITANSDBTHLLOGQSLTTLTLESPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG	180				
12	LVFGITANSDBTHLLOGQSLTTLTLESPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG	180				
13	LVFGITANSDBTHLLOGQSLTTLTLESPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG	180				
14	LVFGITANSDBTHLLOGQSLTTLTLESPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG	180				
15	LVFGITANSDBTHLLOGQSLTTLTLESPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG	180				
16	LVFGITANSDBTHLLOGQSLTTLTLESPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG	180				
17	LVFGITANSDBTHLLOGQSLTTLTLESPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG	180				
18	LVFGITANSDBTHLLOGQSLTTLTLESPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG	180				
19	LVFGITANSDBTHLLOGQSLTTLTLESPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG	180				
20	LVFGITANSDBTHLLOGQSLTTLTLESPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG	180				
21	LVFGITANSDBTHLLOGQSLTTLTLESPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG	180				
22	LVFGITANSDBTHLLOGQSLTTLTLESPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG	180				
23	LVFGITANSDBTHLLOGQSLTTLTLESPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG	180				
24	LVFGITANSDBTHLLOGQSLTTLTLESPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG	180				
25	LVFGITANSDBTHLLOGQSLTTLTLESPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG	180				
26	LVFGITANSDBTHLLOGQSLTTLTLESPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG	180				
27	LVFGITANSDBTHLLOGQSLTTLTLESPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG	180				
28	LVFGITANSDBTHLLOGQSLTTLTLESPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG	180				
29	LVFGITANSDBTHLLOGQSLTTLTLESPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG	180				
30	LVFGITANSDBTHLLOGQSLTTLTLESPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG	180				
31	LVFGITANSDBTHLLOGQSLTTLTLESPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG	180				
32	LVFGITANSDBTHLLOGQSLTTLTLESPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG	180				
33	LVFGITANSDBTHLLOGQSLTTLTLESPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG	180				
34	LVFGITANSDBTHLLOGQSLTTLTLESPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG	180				
35	LVFGITANSDBTHLLOGQSLTTLTLESPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG	180				
36	LVFGITANSDBTHLLOGQSLTTLTLESPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG	180				
37	LVFGITANSDBTHLLOGQSLTTLTLESPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG	180				
38	LVFGITANSDBTHLLOGQSLTTLTLESPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG	180				
39	LVFGITANSDBTHLLOGQSLTTLTLESPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG	180				
40	LVFGITANSDBTHLLOGQSLTTLTLESPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG	180				
41	LVFGITANSDBTHLLOGQSLTTLTLESPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG	180				
42	LVFGITANSDBTHLLOGQSLTTLTLESPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG	180				
43	LVFGITANSDBTHLLOGQSLTTLTLESPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG	180				
44	LVFGITANSDBTHLLOGQSLTTLTLESPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG	180				
45	LVFGITANSDBTHLLOGQSLTTLTLESPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG	180				
46	LVFGITANSDBTHLLOGQSLTTLTLESPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG	180				
47	LVFGITANSDBTHLLOGQSLTTLTLESPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG	180				
48	LVFGITANSDBTHLLOGQSLTTLTLESPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG	180				
49	LVFGITANSDBTHLLOGQSLTTLTLESPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG	180				
50	LVFGITANSDBTHLLOGQSLTTLTLESPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG	180				
51						

[illegible]

```

QY 24 QGNKVVLAGKKGDVLTCTASQKKSIOFHKNKSNQIKIINGQSSFLTKGSPSKLNDRAISR 83
DB 1 QGNKVVLAGKKGDVLTCTASQKKSIOFHKNKSNQIKIINGQSSFLTKGSPSKLNDRAISR 60
QY 84 RSLMDQGNFPLIINKLIKIEDSDTYICEVEDQKEEVLVFGLTANSPTHLLOGQSITLTL 143
DB 61 RSLMDQGNFPLIINKLIKIEDSDTYICEVEDQKEEVLVFGLTANSPTHLLOGQSITLTL 120
QY 144 ESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDSDGTWCTVLOQKKEFKIDIVLVA 203
DB 121 ESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDSDGTWCTVLOQKKEFKIDIVLVA 180
QY 204 FQKASSIVYKKEGEQVEFSPFLAFTVEKLTGSGELMWOABRASSSSKSWITFDLKNKEVS 263
DB 181 FQKASSIVYKKEGEQVEFSPFLAFTVEKLTGSGELMWOABRASSSSKSWITFDLKNKEVS 240
QY 264 KRVTDPKLOMGKKLPLHLTLPQALPOYAGSGNLTALBAKTKLHOEVNLVVMRATOLQ 323
DB 241 KRVTDPKLOMGKKLPLHLTLPQALPOYAGSGNLTALBAKTKLHOEVNLVVMRATOLQ 300
QY 324 KNITCEVWGPTSPKMLSLKENKEAKVSKREKPVWVTLNPEAGMOCCLSDSGQVLLSEN 383
DB 301 KNITCEVWGPTSPKMLSLKENKEAKVSKREKPVWVTLNPEAGMOCCLSDSGQVLLSEN 360
QY 384 IKVLPWTSTVPKPPAPKSCDKTHTCPELLGGPSVFLFPKPKPDITMISRTPEVTCVVY 443
DB 361 IKVLPWTSTVPKPPAPKSCDKTHTCPELLGGPSVFLFPKPKPDITMISRTPEVTCVVY 406
QY 444 DVSHEDPEVKFNWYV-----DGVVHNAAKTK 469
DB 407 REYAADPEYRKHLIEFYKILTNTPDNDEVERRNA DNK 443

RESULT 89
AAP90833
ID AAP90833 standard; protein; 384 AA.
XX
AC AAP90833;
XX
DT 25-MAR-2003 (revised)
DT 01-AUG-1990 (first entry)
XX
DE Amino acid sequence of a soluble T4-like (ST4) polypeptide encoded by a
DE portion of clone p199-7 (PL mutet.rst4).
XX
KW HIV; immunotherapeutic; prophylactic; soluble T4-like polypeptide;
KW diagnostic; p199-7 (PL mutet.rst4).
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 10 /note= "initiating Met"
FT
XX
XX W08901940-A.
XX
XX PD 09-MAR-1989.
XX
XX PF 01-SEP-1988; 88WO-US002940.
XX
XX PR 04-SEP-1987; 87US-00094322.
XX
XX PR 07-JAN-1988; 88US-00141649.
XX
XX PA (BIOI ) BIOGEN INC.
XX
XX PI Fisher RA, Gilbert W, Sato VL, Flavell RA, Maraganore JM;
XX
XX DR WPI: 1989-085519/11.
XX
XX DR N-PSDB; AAN90643.
XX
XX PT DNA sequences coding for soluble T4-like polypeptide(s) - used in
XX immunotherapeutic and immunosuppressive compans. and for preventing,
XX treating or detecting AIDS.

```

```

XX
PS Disclosure; Page 7; 207pp; English.
XX
CC It is the protein sequence encoded by the rst4 sequence. It is claimed in
CC the patent. It is useful in immunotherapeutic, prophylactic and
CC diagnostic compans. It can be used to purify HIV from a sample. (updated
CC on 25-MAR-2003 to correct PR field.)
XX
SQ Sequence 384 AA;
Query Match 55.9%; Score 1908; DB 1; Length 384;
Best Local Similarity 99.7%; Pred. No. 3,4e-95;
Matches 370; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 24 QGNKVVLAGKKGDVLTCTASQKKSIOFHKNKSNQIKIINGQSSFLTKGSPSKLNDRAISR 83
DB 11 QGNKVVLAGKKGDVLTCTASQKKSIOFHKNKSNQIKIINGQSSFLTKGSPSKLNDRAISR 70
QY 84 RSLMDQGNFPLIINKLIKIEDSDTYICEVEDQKEEVLVFGLTANSPTHLLOGQSITLTL 143
DB 71 RSLMDQGNFPLIINKLIKIEDSDTYICEVEDQKEEVLVFGLTANSPTHLLOGQSITLTL 130
QY 144 ESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDSDGTWCTVLOQKKEFKIDIVLVA 203
DB 131 ESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDSDGTWCTVLOQKKEFKIDIVLVA 190
QY 204 FQKASSIVYKKEGEQVEFSPFLAFTVEKLTGSGELMWOABRASSSSKSWITFDLKNKEVS 263
DB 191 FQKASSIVYKKEGEQVEFSPFLAFTVEKLTGSGELMWOABRASSSSKSWITFDLKNKEVS 250
QY 264 KRVTDPKLOMGKKLPLHLTLPQALPOYAGSGNLTALBAKTKLHOEVNLVVMRATOLQ 323
DB 251 KRVTDPKLOMGKKLPLHLTLPQALPOYAGSGNLTALBAKTKLHOEVNLVVMRATOLQ 310
QY 324 KNITCEVWGPTSPKMLSLKENKEAKVSKREKPVWVTLNPEAGMOCCLSDSGQVLLSEN 383
DB 311 KNITCEVWGPTSPKMLSLKENKEAKVSKREKPVWVTLNPEAGMOCCLSDSGQVLLSEN 370
QY 384 IKVLPWTSTVP 394
DB 371 IKVLPWTSTVP 381

RESULT 90
AAR41043
ID AAR41043 standard; protein; 1786 AA.
XX
AC AAR41043;
XX
DT 24-OCT-2003 (revised)
DT 25-MAR-2003 (revised)
DT 22-MAR-1994 (first entry)
XX
DE CD4-EBA175 fusion protein.
XX
XX KW Merozoite; Erythrocyte Binding Antigen 175; malaria; HIV; env;
KW human immunodeficiency virus; envelope glycoprotein; hybrid protein;
KW red blood cell; erythrocyte; AIDS; molecular machine.
XX
XX OS Homo sapiens.
XX
XX OS Plasmodium falciparum.
XX
XX OS Chimeric.
XX
FH Key Location/Qualifiers
FT Region 1..371
FT Region /note= "residues 1-371 of CD4"
FT Region 372..1786
FT Region /note= "residues 20-1435 of EBA-175"
XX
XX W09318160-A1.
XX
XX PD 16-SEP-1993.
XX

```

```

PF 10-MAR-1993; 93WO-GB000505.
XX
PR 11-MAR-1992; 92GB-00005276.
PR 08-JUL-1992; 92GB-00014481.
PR 24-JUL-1992; 92GB-00015829.
PR 16-SEP-1992; 92GB-00019562.
PR 03-MAR-1993; 93GB-00004311.
XX
PA (PREN/) PRENDERGAST K F.
XX
PI Prendergast KF;
XX
DR WPI; 1993-303474/38.
XX
PT Anti-viral fusion peptide(s) - comprise viral-binding component and
PT malaria merozoite red cell binding component, for treating e.g. HIV, and
PT hepatitis.
XX
PS Claim 9; Page 44-47; 69pp; English.
XX
CC The hybrid protein NH2-CD4(1-371)-EBA175(20-1435)-COOH is a specifically
CC claimed example of a fusion protein of the invention; it comprises at
CC least part of the CD4 molecule fused to a peptide from a malarial
CC parasite merozoite protein with affinity for red blood cells. The fusion
CC protein can bind free HIV in the blood to red blood cells and
CC consequently reduce viral titre, prevent transmission of the virus and
CC improve safety of blood transfusions. (Updated on 25-MAR-2003 to correct
CC PN field.) (Updated on 24-OCT-2003 to standardise OS field)
XX
SQ Sequence 1786 AA;
XX
Query Match 55.9%; Score 1908; DB 2; Length 1786;
Best Local Similarity 99.7%; Pred. No. 1.7e-94;
Matches 370; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 24 QGNKVLGKGGDTVELTCTASQKKSIOFHMKNQIKILGNQGSFLTKGPKLNDRADSR 83
DB 1 QGNKVLGKGGDTVELTCTASQKKSIOFHMKNQIKILGNQGSFLTKGPKLNDRADSR 60
QY 84 RSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEVQLVFGLTANSPTHLLOGSLTTL 143
DB 61 RSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEVQLVFGLTANSPTHLLOGSLTTL 120
QY 144 ESPGSSPVQCRSPRGKNIQGGKTLVSQLELDSDGTWCTVYLQNKVKEFKIDIVVLA 203
DB 121 ESPGSSPVQCRSPRGKNIQGGKTLVSQLELDSDGTWCTVYLQNKVKEFKIDIVVLA 180
QY 204 FOKASSIIVYKKEGEQVFFSPPLAFTVEKLTGSGELMQAERASSKSWITFDLKNKEYSV 263
DB 181 FOKASSIIVYKKEGEQVFFSPPLAFTVEKLTGSGELMQAERASSKSWITFDLKNKEYSV 240
QY 264 KRYTQDPKLGKGLPLHLTLPLALPOYAGSGNLTALAEAKTGKLGHEVNLVVMRATQLQ 323
DB 241 KRYTQDPKLGKGLPLHLTLPLALPOYAGSGNLTALAEAKTGKLGHEVNLVVMRATQLQ 300
QY 324 KNLTCEVWGPTSPKLMSTLKENKEAKVSKREKPVWVNLNPAAGMOCILSDSGVLLSEN 383
DB 301 KNLTCEVWGPTSPKLMSTLKENKEAKVSKREKPVWVNLNPAAGMOCILSDSGVLLSEN 360
QY 384 IKVLPWTWSTPV 394
DB 361 IKVLPWTWSTPV 371
XX
RESULT 91
ID AAP96151 standard; protein; 434 AA.
AC AAP96151;
XX
XX 25-MAR-2003 (revised)
DT 24-DEC-1990 (first entry)
XX

```

```

DE Sequence of a fusion of the herpes virus and N-terminal 27 residues to
DE the putative mature N-terminus of CD4T.
XX
KM gp120; HIV; therapy; adhesion; antiviral.
XX
OS Synthetic.
XX
PN EP14317-A.
XX
PD 03-MAY-1989.
XX
PP 03-OCT-1988; 88EP-00309194.
XX
PR 02-OCT-1987; 87US-00104329.
PR 28-SEP-1988; 88US-00250785.
XX
PA (GETH ) GENENTECH INC.
XX
PI Capon DJ, Gregory TJ;
XX
DR WPI; 1989-131855/18.
DR N-PSDB; AAN90778.
XX
CC Compns. contg. adhesion variants - useful in therapy and diagnostics,
PT e.g. CD4 variants which are therapeutically useful for treating human
PT immuno-deficiency virus.
XX
PS Example; Fig 2a-2c; 36pp; English.
XX
CC A nucleic acid encoding an AA sequence variant of an adhesion is claimed.
CC Adhesion variants may be a fusion of a CD4 polypeptide and a polypeptide
CC different from CD4. Used therapeutically or as diagnostic reagents for
CC the assay of adhesions or their ligands, or for purificn. of the adhesions.
CC The CD4 adhesion variants are used for antiviral or immunomodulatory
CC therapy, in partic. treatment of HIV infection. (Updated on 25-MAR-2003
CC to correct PI field.)
XX
SQ Sequence 434 AA;
XX
Query Match 55.8%; Score 1904; DB 1; Length 434;
Best Local Similarity 99.7%; Pred. No. 6.3e-95;
Matches 369; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 24 QGNKVLGKGGDTVELTCTASQKKSIOFHMKNQIKILGNQGSFLTKGPKLNDRADSR 83
DB 56 QGNKVLGKGGDTVELTCTASQKKSIOFHMKNQIKILGNQGSFLTKGPKLNDRADSR 115
QY 84 RSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEVQLVFGLTANSPTHLLOGSLTTL 143
DB 116 RSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEVQLVFGLTANSPTHLLOGSLTTL 175
QY 144 ESPGSSPVQCRSPRGKNIQGGKTLVSQLELDSDGTWCTVYLQNKVKEFKIDIVVLA 203
DB 176 ESPGSSPVQCRSPRGKNIQGGKTLVSQLELDSDGTWCTVYLQNKVKEFKIDIVVLA 235
QY 204 FOKASSIIVYKKEGEQVFFSPPLAFTVEKLTGSGELMQAERASSKSWITFDLKNKEYSV 263
DB 236 FOKASSIIVYKKEGEQVFFSPPLAFTVEKLTGSGELMQAERASSKSWITFDLKNKEYSV 295
QY 264 KRYTQDPKLGKGLPLHLTLPLALPOYAGSGNLTALAEAKTGKLGHEVNLVVMRATQLQ 323
DB 296 KRYTQDPKLGKGLPLHLTLPLALPOYAGSGNLTALAEAKTGKLGHEVNLVVMRATQLQ 355
QY 324 KNLTCEVWGPTSPKLMSTLKENKEAKVSKREKPVWVNLNPAAGMOCILSDSGVLLSEN 383
DB 356 KNLTCEVWGPTSPKLMSTLKENKEAKVSKREKPVWVNLNPAAGMOCILSDSGVLLSEN 415
QY 384 IKVLPWTWSTP 393
DB 416 IKVLPWTWSTP 425
XX
RESULT 92

```

AA093557
ID AA093557 standard; protein; 434 AA.
AC AAP93557;
XX
XX 25-MAR-2003 (revised)
DT 03-OCT-2002 (revised)
DT 06-JUN-1990 (first entry)
XX
XX Fusion of the herpes gD leader and N-terminal 27 residues to mature N-terminus of CD4T.
DE
XX Adhesion variant; CD4T; soluble CD4; truncated CD4; antiviral; HIV;
KM herpes gD; immunomodulatory; diagnostic.
XX
XX Homo sapiens.
OS
XX MO8902922-A.
PN
XX 06-APR-1989.
PD
XX 03-OCT-1988; 88WO-US003414.
PF
XX 02-OCT-1987; 87US-00104329.
PR 28-SEP-1988; 88US-00250785.
XX
XX (GETH) GENENTECH INC.
PA
XX Capon DJ, Gregory TJ;
PI
XX WPI; 1989-114397/15.
DR N-PSDB; AAN90735.
XX
XX New nucleic acid sequences encoding adhesion, esp. CD 4, variants -
PT partic. with trans-membrane domain inactivated or fused to other peptide,
PT useful esp. for treating HIV infections.
XX
XX Fig 2A-2C; pp. 5/13-8/13; 78pp; English.
PS
XX CD4T is a truncated or soluble variant of CD4. CD4T fusion proteins can
CC have antiviral and immunomodulatory activity and are esp. useful for
CC treating HIV infections, regardless of genetic variation within the
CC virus. CD4T fusion proteins, and antibodies raised against them, can also
CC be used diagnostically for assaying adhesions and their ligands. (Updated
CC on 03-OCT-2002 to add missing OS field.) (Updated on 25-MAR-2003 to
CC correct PR field.) (Updated on 25-MAR-2003 to correct PA field.)
XX
XX
SQ Sequence 434 AA;
Query Match 55.8%; Score 1904; DB 1; Length 434;
Best Local Similarity 99.7%; Pred. No. 6.3e-95;
Matches 369; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 324 KNLTCFVWGPTSPKMLSLKLENKEAKVSKREKPVVILNPBAQMOCLLSDSGQVLLSESN 383
DB |||||
DB 356 KNLTCFVWGPTSPKMLSLKLENKEAKVSKREKAVVILNPBAQMOCLLSDSGQVLLSESN 415
QY 384 IKVLPWSTP 393
DB |||||
DB 416 IKVLPWSTP 425
RESULT 93
AA07721
ID AA07721 standard; protein; 375 AA.
XX
XX AA07721;
AC
XX 25-MAR-2003 (revised)
DT 18-FEB-1991 (first entry)
XX
XX Recombinant soluble (rs) T4.
DE
XX
XX Metal-binding site.
KM
XX Homo sapiens.
OS
XX MO9012803-A.
PN
XX 01-NOV-1990.
PD
XX 14-APR-1989; 89US-00338991.
PF
XX 14-APR-1989; 89US-00338991.
PR 14-APR-1989; 89US-00338991.
XX
XX (BIOI) BIOGEN INC.
PA (BIOI) BIOGEN INC.
XX
XX Staples MA, Pargellis CA;
PI
XX WPI; 1990-348421/46.
DR
XX Purifying protein having surface metal-binding amino acid residues -
PT using an immobilised metal affinity chromatography resin.
PT
XX
XX Disclosure; Fig 2; 36pp; E.
PS
XX
XX The rs T4 can be purified from a crude sample (Updated on 25-MAR-2003 to
CC correct PA field.)
CC
XX
XX
SQ Sequence 375 AA;
Query Match 55.5%; Score 1894; DB 2; Length 375;
Best Local Similarity 99.5%; Pred. No. 1.9e-94;
Matches 367; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 326 LTCEVWGPTSPKMLSLKLENKEAKVSKREKRPVWVNLPEAGMOCCLSDSGQVLLSENIK 385
 Db 301 LTCEVWGPTSPKMLSLKLENKEAKVSKREKRPVWVNLPEAGMOCCLSDSGQVLLSENIK 360
 Qy 386 VLPTWSTPV 394
 Db 361 VLPTWSTPV 369

RESULT 94

AAW41376
 ID AAW41376 standard; peptide; 433 AA.

AC AAW41376;

DT 28-MAY-1998 (first entry)

DE Human CD4.

KM Antibody; CD4; passive immunity; HIV type 1; HIV type 2; HIV infection;
 KM simian immunodeficiency virus; SIV; AIDS; therapy; HIV gp120.

OS Homo sapiens.

PN MO9746697-A2.

PD 11-DEC-1997.

PF 03-JUN-1997; 97WO-US009449.

PR 03-JUN-1996; 96US-00657149.

PR 28-FEB-1997; 97US-00808374.

PR 02-JUN-1997; 97US-00867149.

PA (UNBI-) UNITED BIOMEDICAL INC.

PI Wang CY;

DR WPI: 1998-042204/04.

XX Antibody against complex of CD4 and chemokine receptor domain - useful
 XX for prevention and treatment of human immunodeficiency virus infection.

XX Disclosure: Page 122-123; 140pp; English.

XX This sequence represents human CD4, and was used to isolate the antibody
 XX (Ab) of the invention. The Ab, preferably a M2 or B13 Ab, has the
 XX following characteristics: (a) binds rscD4; (b) binds CD4 expressing
 XX cells in an immunofluorescence assay, where the binding pattern is in the
 XX shape of "caps", when examined with a high resolution fluorescence
 XX microscope; (c) blocks the binding of HIV glycoprotein 120 (gp120) to CD4
 XX expressing cells; (d) binds CD4 expressing cells previously bound with
 XX gp120; and (e) neutralises HIV primary isolates in an in vitro
 XX microplaque assay at a concentration of less than 10 mu g/ml. Preferably
 XX at a concentration in the range of 0.01-10 mu g/ml for 50% neutralisation
 XX and 0.1-35 mu g/ml for 90% neutralisation. The Ab can be used to provide
 XX passive immunity to HIV in a mammal, when administered parenterally,
 XX specifically all clades of HIV type 1, and from diverse primary isolates
 XX of HIV type 2 and simian immunodeficiency virus (SIV). The Ab is
 XX prophylactic and therapeutic for HIV infection and all stages of AIDS
 XX because it prevents replicative infection of host cells both before and
 XX after HIV gp120 has bound to the host cell antigen complex comprising CD4
 XX on the surface of CD4 positive lymphocytes, thus it is capable of
 XX preventing HIV infection and retarding the spread of the virus to
 XX uninfected cells. It is also uniquely useful because it inhibits
 XX infection following binding of HIV to CD4 expressing cells

XX Sequence 433 AA;

Query Match 55.4%; Score 1892; DB 2; Length 433;
 Best Local Similarity 99.5%; Pred. No. 2.8e-94;
 Matches 367; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 26 NKVVLGKKGDVVELTCTASQKSIQFHMKNNSNOIKILGNQSPFLTKGPSKLNDRADSRRS 85
 Db 1 NKVVLGKKGDVVELTCTASQKSIQFHMKNNSNOIKILGNQSPFLTKGPSKLNDRADSRRS 60
 Qy 86 LMDQGNFPLIIKNLIKIDSDTYICEVEDQKEVQVLVFGITANSDFHLQGSITLTLES 145
 Db 61 LMDQGNFPLIIKNLIKIDSDTYICEVEDQKEVQVLVFGITANSDFHLQGSITLTLES 120
 Qy 146 PGSSPSVQCRSPRGKNIQGGKTLTSSQLELODSGTCTVLONOKVEFFIDIVLAFQ 205
 Db 121 PGSSPSVQCRSPRGKNIQGGKTLTSSQLELODSGTCTVLONOKVEFFIDIVLAFQ 180
 Qy 206 KASSIVYKKEGEQVEFSPLAFTEVEKLTSGGELMWQERASSSKSWITPDLKNEVSVKR 265
 Db 181 KASSIVYKKEGEQVEFSPLAFTEVEKLTSGGELMWQERASSSKSWITPDLKNEVSVKR 240
 Qy 266 VTQDPKIQMGKPLHLTLFQALPQVAGSGLTLALEAKTGKQHVEVNLVVMRATQIQKN 325
 Db 241 VTQDPKIQMGKPLHLTLFQALPQVAGSGLTLALEAKTGKQHVEVNLVVMRATQIQKN 300
 Qy 326 LTCEVWGPTSPKMLSLKLENKEAKVSKREKRPVWVNLPEAGMOCCLSDSGQVLLSENIK 385
 Db 301 LTCEVWGPTSPKMLSLKLENKEAKVSKREKRPVWVNLPEAGMOCCLSDSGQVLLSENIK 360
 Qy 386 VLPTWSTPV 394
 Db 361 VLPTWSTPV 369

RESULT 95

AAV39824
 ID AAV39824 standard; protein; 369 AA.

AC AAV39824;

DT 03-DEC-1999 (first entry)

DE Soluble human T4 protein.

KM Soluble T4 protein; sT4; human; HIV; binding inhibitor; T4+ cell; AIDS;
 KM vaccine; immunisation; therapy.

OS Homo sapiens.

PN US5958678-A.

PD 28-SEP-1999.

PF 12-DEC-1994; 94US-00354452.

PR 21-AUG-1986; 86US-00898587.

PR 11-JUN-1991; 91US-00713564.

PR 06-JUL-1992; 92US-00909021.

PA (UYCO) UNIV COLUMBIA NEW YORK.

PI McDougall JS, Weiss R, Axel R, Littman DR, Maddon PJ, Chess L;

DR WPI: 1999-561025/47.

XX Human T4 protein inhibits HIV binding to T4 cells, useful for treating
 XX AIDS.

XX Claim 1; Col 51-53; 58pp; English.

XX This sequence represents the soluble human T4 protein of the invention.
 XX The soluble human T4 protein blocks the binding of HIV to T4+ cells and
 XX is therefore useful for the treatment of AIDS. Monoclonal antibodies
 XX against the T4 protein may be used as vaccines for immunising subjects
 XX against AIDS

XX Sequence 369 AA;

Query Match 55.4%; Score 1891; DB 2; Length 369;
 Best Local Similarity 99.7%; Pred. No. 2.7e-94;
 Matches 367; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 27 KVLGKKGDVLTCTASQKSIQFHWKNSNQIKILNGSFLTKGPSKLNDRADSRSL 86
 CC 2 KVLGKKGDVLTCTASQKSIQFHWKNSNQIKILNGSFLTKGPSKLNDRADSRSL 61
 DB 87 WDOGNFPLIKNLKIEDSDTYICEVEDQKEVOLVFGLTANSDFHLLOGSLTTLTLESP 146
 DB 62 WDOGNFPLIKNLKIEDSDTYICEVEDQKEVOLVFGLTANSDFHLLOGSLTTLTLESP 121
 QY 147 PGSSPSVQCRSPRKNIOGKTLVSQLELDQSGTWCTVLONOKKVEFKIDIVLAFOK 206
 DB 122 PGSSPSVQCRSPRKNIOGKTLVSQLELDQSGTWCTVLONOKKVEFKIDIVLAFOK 181
 QY 207 ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELMWQERASSSKSWITFDLKNKEVSVKRV 266
 DB 182 ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELMWQERASSSKSWITFDLKNKEVSVKRV 241
 QY 267 TODPKLQMGKKLPLHLTLTPQALPOYAGSGNLTALAEAKTGKLEHENVLVVNRATQLOKRL 326
 DB 242 TODPKLQMGKKLPLHLTLTPQALPOYAGSGNLTALAEAKTGKLEHENVLVVNRATQLOKRL 301
 QY 327 TCEVWGPTSPKMLSLKLENKEAKVSKREKPVWVNLPEAGMOCCLSDSGOVLLESNIRY 386
 DB 302 TCEVWGPTSPKMLSLKLENKEAKVSKREKPVWVNLPEAGMOCCLSDSGOVLLESNIRY 361
 QY 387 LPTWSTPV 394
 DB 362 LPTWSTPV 369

RESULT 96
 AAY88327
 ID AAY88327 standard; protein; 369 AA.
 AC AAY88327;
 XX 14-JUL-2000 (first entry)
 DT 14-JUL-2000 (first entry)
 DE T4 glycoprotein amino acid sequence.
 XX
 KM sT4; glycoprotein; human immunodeficiency virus; HIV; block binding;
 KM AIDS; treatment; inhibit; cell to cell spread; infection; fusion.
 XX
 OS Mammalia.
 XX
 PN USS126433-A.
 XX 30-JUN-1992.
 PD 30-JUN-1992.
 PF 23-OCT-1987; 87US-00114244.
 XX 21-AUG-1986; 86US-00898587.
 PR 21-AUG-1986; 86US-00898587.
 XX (UYCO) UNIV COLUMBIA NEW YORK.
 PA
 PI Maddon PJ, Chess L, Axel R, Weiss R, Littman DR, McDougal JS;
 XX WPI; 2000-348913/30.
 DR
 XX Soluble T4 glycoprotein useful for prevention and treatment of acquired
 PT immunodeficiency syndrome and for screening inhibitors of human
 PT immunodeficiency viral binding.
 XX
 PS Claim 1; Col 54; 64pp; English.
 CC This sequence represents the amino acid sequence of glycosylated sT4
 CC glycoprotein. Human immunodeficiency virus (HIV) uses sT4 as a target
 CC receptor on T cells. The invention relates to glycosylated sT4 which
 CC functions by blocking the binding of HIV to T4 target cells, and can be
 CC used for the prophylaxis and treatment of AIDS patients. Administration

CC of sT4 effectively inhibits the cell to cell spreading of HIV infection
 CC and also the fusion of HIV-infected T4 cells and non-infected T4 cells.
 CC The administration of T4 alleviates several symptoms associated with
 CC AIDS, and prevents the occurrence of new pathological changes. The sT4
 CC glycoprotein is useful for the prophylaxis and treatment of patients with
 CC AIDS. It is also useful as a reagent to identify natural, synthetic or
 CC recombinant molecules which act as therapeutic agents or inhibitors of
 CC T4+ cell interactions and in diagnostic assays for detection T4 proteins
 CC or molecules

Sequence 369 AA;

QY 27 KVLGKKGDVLTCTASQKSIQFHWKNSNQIKILNGSFLTKGPSKLNDRADSRSL 86
 DB 2 KVLGKKGDVLTCTASQKSIQFHWKNSNQIKILNGSFLTKGPSKLNDRADSRSL 61
 QY 87 WDOGNFPLIKNLKIEDSDTYICEVEDQKEVOLVFGLTANSDFHLLOGSLTTLTLESP 146
 DB 62 WDOGNFPLIKNLKIEDSDTYICEVEDQKEVOLVFGLTANSDFHLLOGSLTTLTLESP 121
 QY 147 PGSSPSVQCRSPRKNIOGKTLVSQLELDQSGTWCTVLONOKKVEFKIDIVLAFOK 206
 DB 122 PGSSPSVQCRSPRKNIOGKTLVSQLELDQSGTWCTVLONOKKVEFKIDIVLAFOK 181
 QY 207 ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELMWQERASSSKSWITFDLKNKEVSVKRV 266
 DB 182 ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELMWQERASSSKSWITFDLKNKEVSVKRV 241
 QY 267 TODPKLQMGKKLPLHLTLTPQALPOYAGSGNLTALAEAKTGKLEHENVLVVNRATQLOKRL 326
 DB 242 TODPKLQMGKKLPLHLTLTPQALPOYAGSGNLTALAEAKTGKLEHENVLVVNRATQLOKRL 301
 QY 327 TCEVWGPTSPKMLSLKLENKEAKVSKREKPVWVNLPEAGMOCCLSDSGOVLLESNIRY 386
 DB 302 TCEVWGPTSPKMLSLKLENKEAKVSKREKPVWVNLPEAGMOCCLSDSGOVLLESNIRY 361
 QY 387 LPTWSTPV 394
 DB 362 LPTWSTPV 369

RESULT 97
 AAP93528
 ID AAP93528 standard; protein; 370 AA.
 AC AAP93528;
 XX 25-MAR-2003 (revised)
 DT 04-JUN-1990 (first entry)
 DE Human soluble CD4 protein (T4ex1) encoded by T4 SEC1 cDNA.
 XX
 KM Human soluble CD4 protein; T4 SEC1 cDNA; T4ex1;
 KM HIV gp120 envelope protein; T-lymphocyte.
 XX
 OS Homo sapiens.
 XX
 PN Key Location/Qualifiers
 FH Key Location/Qualifiers
 FT Misc-difference 129
 FT Misc-difference 163
 FT Misc-difference 163
 FT Misc-difference 243
 FT Misc-difference 370
 XX
 PN W08903222-A.
 XX 20-APR-1989.
 PD 20-APR-1989.
 PF 05-OCT-1988; 88MO-US003454.

```

XX 08-OCT-1987; 87US-00106185.
PR 14-JAN-1988; 88US-00144313.
PR 14-JUN-1988; 88US-00206585.
PR 11-JUL-1988; 88US-00217475.
XX
PA (DANA-) DANA-FABER CANCER.
XX
PI Reinherz E, Huseby R, Droski J, Richardson N;
DR WPI; 1989-122898/17.
DR N-PSDB; AAN90764.
XX
PT Soluble human CD4 fragments - capable of binding HIV gp 120 envelope
PT protein, used for diagnosis, treatment or prophylaxis of HIV infection.
XX
PS Fig 1; pages 1/11-3/11; 106pp; English.
XX
CC Misc-difference in the features table represent where a modification in
CC cDNA and in the encoded CD4 protein has been made. A soluble human CD4
CC fragment capable of binding to the gp. 120 envelope protein of HIV is
CC claimed. Such a fragment should include none of the hydrophobic trans-
CC membrane region of naturally-occurring CD4, or contain a portion of the
CC hydrophobic region which is sufficiently short it does not prevent
CC solubilization. Also claimed are modified soluble CD4 fragments which
CC differ by subseqn. or addn. to this sequence (AAP93528) by at least one
CC AA. Such fragments pref. consist of: the first 182, 369, 549, 729 or 1107
CC AA, domain 1,2, and partial domain 3-pref. where the protein is
CC truncated at AA posn. 243; Asn-271 is Asp and/or Asn-300 is Asp; or the
CC AA sequence without the N-linked glycosylation sites; of CD4 protein.
CC Such fragments can also contain the corresp. AA of murine CD4 subseq. at
CC more than one triplet site in the DNA, selected from a triplet encoding
CC the AA at posn. 48; 50; 51; 121; 122; 123; 155; 156; or 158 of human CD4.
CC CD4 fragments have the capacity to prevent infection of human T-
CC lymphocyte infection by HIV and to prevent the formation of human T-
CC lymphocyte syncytia. They can be used in the diagnosis, treatment and
CC prevention of HIV infection. (Updated on 25-MAR-2003 to correct PR
CC field.) (Updated on 25-MAR-2003 to correct PI field.)
XX
SQ Sequence 370 AA:
Query Match 55.4%; Score 1891; DB 1; Length 370;
Best Local Similarity 99.7%; Pred. No. 2.7e-94;
Matches 367; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 27 KVVVGKKGDVVELTCTASQKKSIOFHWKNSNQIKLGNQGSFLTKGPSKINDRADSRRL 86
DB 2 KVVVGKKGDVVELTCTASQKKSIOFHWKNSNQIKLGNQGSFLTKGPSKINDRADSRRL 61
QY 87 WDOGNFPLITIKNLKIEBSPDYICEVEDQKEEVLVFGLTANSPTHTLLOQGSLLTLTLESP 146
DB 62 WDOGNFPLITIKNLKIEBSPDYICEVEDQKEEVLVFGLTANSPTHTLLOQGSLLTLTLESP 121
QY 147 PGSSPSVQCRSPRGKNIQGGKTLVSQLELDGSGTWTCTVLQNKQKVEFKIDIVLAFOK 206
DB 122 PGSSPSVQCRSPRGKNIQGGKTLVSQLELDGSGTWTCTVLQNKQKVEFKIDIVLAFOK 181
QY 207 ASSIVYKKEGQVEFSPPLAFTVEKLTGSGELMWQAERASSSKSWITFDLKNKEVSYKRV 266
DB 182 ASSIVYKKEGQVEFSPPLAFTVEKLTGSGELMWQAERASSSKSWITFDLKNKEVSYKRV 241
QY 267 TODPKLQMGKPLHLTLPLPOLROYASGNLTALAEKGTGLHDEVNLVYMRARQLOK 326
DB 242 TODPKLQMGKPLHLTLPLPOLROYASGNLTALAEKGTGLHDEVNLVYMRARQLOK 301
QY 327 TCEVWGPTSPKMLSLKLENKEAKVSKREKPVVYLNPEAGMOCCLSDSGCVLLSNIKY 386
DB 302 TCEVWGPTSPKMLSLKLENKEAKVSKREKPVVYLNPEAGMOCCLSDSGCVLLSNIKY 361
QY 387 LPTWSTPV 394
DB 362 LPTWSTPV 369

```

```

RESULT 98
AAB83356
ID AAB83356 standard; protein; 370 AA.
XX
XX AAB83356;
AC
XX
DT 09-OCT-2001 (first entry)
XX
DE Human CD4 protein sequence.
XX
XX Chemotactic chemokine receptor 5; gp120; CD4; therapy; HIV; CCR5;
XX human immunodeficiency virus; anti-inflammatory disease; human.
XX
OS Homo sapiens.
XX
XX EPI118858-A2.
XX
XX 25-JUL-2001.
XX
XX 03-JAN-2001; 2001EP-00300020.
XX
XX 12-JAN-2000; 2000GB-00000659.
XX 12-JAN-2000; 2000GB-00000661.
XX 12-JAN-2000; 2000GB-00000663.
XX
XX (PFIZ ) PFIZER LTD.
XX (PFIZ ) PFIZER INC.
XX
XX Dobbs S, Perros M, Rickett GA;
XX
XX WPI; 2001-477088/52.
XX
XX N-PSDB; AAP87101.
XX
XX Determining if an agent can modulate CCR5-gp120 interaction, comprises
XX incubating the agent with CCR5 and gp120 and determining if the agent
XX modulates the interaction.
XX
XX Claim 1; Page 112; 113pp; English.
XX
XX
XX This sequence represents the human CD4 protein sequence. The invention
XX relates to a method for determining whether an agent is capable of
XX modulating the interaction of chemotactic chemokine receptor 5 (CCR5)
XX with gp120, comprising incubating the agent with CCR5 and gp120 and
XX determining whether the agent modulates the interaction, where gp120 is
XX associated with CD4, and where the interaction is a low affinity binding.
XX The method is used to identify an agent capable of modulating the
XX interaction of CCR5 with gp120. An agent identified by the method is used
XX to prepare a pharmaceutical composition for the treatment of a disease or
XX condition associated with CCR5 and gp120 interaction, to treat a subject
XX with a disease or condition associated with CCR5 and gp120 interaction,
XX and for preparing a pharmaceutical for treating human immunodeficiency
XX virus (HIV). It can also be used to treat anti-inflammatory diseases. The method
XX is commercially useful, amenable to high throughput screening, and
XX detects interaction of gp120 with cells expressing only CCR5
XX
SQ Sequence 370 AA:
Query Match 55.4%; Score 1891; DB 4; Length 370;
Best Local Similarity 99.7%; Pred. No. 2.7e-94;
Matches 367; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 27 KVVVGKKGDVVELTCTASQKKSIOFHWKNSNQIKLGNQGSFLTKGPSKINDRADSRRL 86
DB 2 KVVVGKKGDVVELTCTASQKKSIOFHWKNSNQIKLGNQGSFLTKGPSKINDRADSRRL 61
QY 87 WDOGNFPLITIKNLKIEBSPDYICEVEDQKEEVLVFGLTANSPTHTLLOQGSLLTLTLESP 146
DB 62 WDOGNFPLITIKNLKIEBSPDYICEVEDQKEEVLVFGLTANSPTHTLLOQGSLLTLTLESP 121
QY 147 PGSSPSVQCRSPRGKNIQGGKTLVSQLELDGSGTWTCTVLQNKQKVEFKIDIVLAFOK 206
DB 122 PGSSPSVQCRSPRGKNIQGGKTLVSQLELDGSGTWTCTVLQNKQKVEFKIDIVLAFOK 181

```

Qy	20	ASSIVYKKDEQVPEFPEPLAFYVTEKLTGSGELIMQAEASSKSMITPDLNKKKVSXRV	266
Db	182	ASSIVYKKDEQVPEFPEPLAFYVTEKLTGSGELIMQAEASSKSMITPDLNKKKVSXRV	244
Qy	267	TDDPKLQMGKKLPLHLTLTPQALPOVAGSGLTLALEAKTKGLHDEVNLVVMRATQLQKNTL	326
Db	242	TDDPKLQMGKKLPLHLTLTPQALPOVAGSGLTLALEAKTKGLHDEVNLVVMRATQLQKNTL	301
Qy	327	TCEVWGPTSPKMLSLKLENKAKNYSKRKKPVMVLPNPEAGMOCILSSGGVLLSEINIKV	386
Db	302	TCEVWGPTSPKMLSLKLENKAKNYSKRKKAVVLPNPEAGMOCILSSGGVLLSEINIKV	361
Qy	387	LPTMSTPV 394	
Db	362	LPTMSTPV 369	
RESULT 99			
AAVS4500			
ID	AAVS4500 standard; protein; 433 AA.		
XX			
AC	AAVS4500;		
XX			
DT	25-APR-2000 (first entry)		
XX			
DE	Amino acid sequence of the human CD4 protein.		
XX			
KW	Human; CD4 protein; antigenic peptide; CDR2-1-like domain; apoptosis;		
KW	syncytia formation; human immune deficiency virus; HIV binding;		
KW	CD4-Class II interaction; immunisation; CD4 surface complex;		
KW	immune response; transplant rejection; autoimmune disease;		
KW	rheumatoid arthritis; systemic lupus erythematosus; protease.		
XX			
OS	Homo sapiens.		
XX			
FH	Key		
FT	Location/Qualifiers		
FT	Domain		
FT	27..66		
FT	note= "CDR-2 like domain; specifically claimed in claim		
FT	1"		
XX			
PN	W09967294-A1.		
XX			
PD	29-DEC-1999.		
XX			
PF	21-JUN-1999; 99MO-US014030.		
XX			
KX	20-JUN-1998; 98US-00100409.		
XX			
PA	(UNBI-) UNITED BIOMEDICAL INC.		
XX			
PI	Wang CY;		
XX			
DR	WPI; 2000-160579/14.		
XX			
PT	New antigenic peptide from the CDR2 domain of CD4, for immunization		
PT	against e.g. human immune deficiency virus.		
XX			
PS	Claim 1; Page 70-71; 106pp; English.		
XX			
CC	The present sequence represents the human CD4 protein. The specification		
CC	describes antigenic peptides derived from the CDR2-like domain of CD4		
CC	(amino acids 27-66 of AAVS4500). These antigenic peptides present		
CC	neutralising receptor/co-receptor effector sites of the CDR2-like domain.		
CC	The peptides evoke effective antibody responses by having optimised site		
CC	specificity. The induced antibodies block human immune deficiency virus		
CC	(HIV) binding and syncytia formation. They may also block CD4-Class II		
CC	interactions with other cells, deliver signals to T cells (inhibiting		
CC	normal CD4+-mediated immunoregulatory functions) or induce apoptosis of		
CC	CD4 cells by simultaneous engagement of T cell receptors. Conjugates and		
CC	peptides containing the antigenic peptides are used for active		
CC	immunisation to generate antibodies against CD4 surface complexes,		
CC	especially to prevent binding of HIV to CD4 and thus HIV infection, but		

Query Match	55.3%	Score 1887	DB 3	Length 433
Best Local Similarity	99.5%	Pred. No. 5.2e-94		
Matches 366	Conservative 0	Mismatches 2	Indels 0	Gaps 0
CC	also to treat undesirable immune responses such as transplant rejection,			
CC	or autoimmune diseases (rheumatoid arthritis, systemic lupus			
CC	erythematosus or psoriasis). These conjugates produce high-titre			
CC	antibodies which are broadly neutralising against primary isolates from			
CC	all classes of HIV-1 and of HIV-2. The peptides may be cyclically			
CC	constrained and may include a promiscuous T helper epitope that is active			
CC	in genetically diverse subjects			
XX				
Sequence 433 AA:				
Query Match	55.3%	Score 1887	DB 3	Length 433
Best Local Similarity	99.5%	Pred. No. 5.2e-94		
Matches 366	Conservative 0	Mismatches 2	Indels 0	Gaps 0
QY	27 KVVLGKKGDTVELTCTTASQKKSIOFHMKNNSQIKILGNQGSFLTKGFSKLNDRADSRSL	86		
DB	2 KVVLGKKGDTVELTCTTASQKKSIOFHMKNNSQIKILGNQGSFLTKGFSKLNDRADSRSL	61		
QY	WDQGNFPLIIKNLKTEDSDPTVCEVEDQKEEVQLVFLGSLTNSDTHLLOGOSLTLTLESP	146		
DB	62 WDQGNFPLIIKNLKTEDSDPTVCEVEDQKEEVQLVFLGSLTNSDTHLLOGOSLTLTLESP	121		
QY	147 PGSSPSPVQCRSPRGKNIQSGKTLVSQELQDSDGTWTCTVQLONQKVEFKIDIVLAFOK	206		
DB	122 PGSSPSPVQCRSPRGKNIQSGKTLVSQELQDSDGTWTCTVQLONQKVEFKIDIVLAFOK	181		
QY	207 ASSIIVYKKEGEQVEFSPLAFTVEKLTGSGELMWOAERASSKSWITFDLKNKEVSVKRV	266		
DB	182 ASSIIVYKKEGEQVEFSPLAFTVEKLTGSGELMWOAERASSKSWITFDLKNKEVSVKRV	241		
QY	267 TDQPKLOMGKPLPHLTLTPOALPOVAGSGNLTALAEATGKLHDEVNLVYMRATQLOKNL	326		
DB	242 TDQPKLOMGKPLPHLTLTPOALPOVAGSGNLTALAEATGKLHDEVNLVYMRATQLOKNL	301		
QY	327 TCEVMGPTSPKMLSLKLENKEAKVSKKEKPVWVNLNPEAGMOCILLSDSGQVLLSNIKV	386		
DB	302 TCEVMGPTSPKMLSLKLENKEAKVSKKEKPVWVNLNPEAGMOCILLSDSGQVLLSNIKV	361		
QY	387 LPTWSTPV 394			
DB	362 LPTWSTPV 369			
RESULT 100				
AAR74222				
ID	AAR74222 standard; protein; 433 AA.			
AC	AAR74222;			
XX				
DT	25-MAR-2003 (revised)			
DT	26-NOV-1995 (first entry)			
XX				
DE	Epitope on the primary CD4 sequence.			
XX				
XX	Chimaeric; mutant; mapping; immunodiagnostics.			
XX	Synthetic.			
XX	OS			
PN	US5411861-A.			
XX				
PD	02-MAY-1995.			
XX				
PF	27-FEB-1992; 92US-00842465.			
XX				
PR	15-APR-1988; 88US-00181826.			
XX				
PA	(GEHO) GEN HOSPITAL CORP.			
XX				
PI	Seed B, Peterson A;			
XX				
DR	WPI; 1995-178122/23.			
XX				
PT	Mutational analysis method for protein epitope(s) - by expressing mutant			

PT cDNA and using negative and positive selection to identify binding loss
PT mutants.

PS Disclosure: Fig 2; 28pp; English.

XX The sequence is that of an epitope on the primary CD4 sequence. The
CC epitope sequence can be identified by vector pIH3M. The See also
CC AAR74221. (Updated on 25-MAR-2003 to correct PF field.)
XX

SO Sequence 432 AA;

Query Match 55.2%; Score 1884; DB 2; Length 432;

Best Local Similarity 99.5%; Pred. No. 7.5e-94;

Matches 366; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY	27	KVLGKKGDTVELTCTAOKKSIOFHMKNNSQIKILGNQGSFLTKGPKLNDRADSRSL	86
DB	1	KVLGKKGDTVELTCTAOKKSIOFHMKNNSQIKILGNQGSFLTKGPKLNDRADSRSL	60
QY	87	WDOGFPILIKNLKIEDSDTYICEVEDQKEEVOLLVFGLTANSPTHLQGSLLTLESF	146
DB	61	WDOGFPILIKNLKIEDSDTYICEVEDQKEEVOLLVFGLTANSPTHLQGSLLTLESF	120
QY	147	PGSSPSVQCRSPRGKNIQGGKTLVSQLELQDSGTWCTVLQNOKVEFKIDIVVLAFOK	206
DB	121	PGSSPSVQCRSPRGKNIQGGKTLVSQLELQDSGTWCTVLQNOKVEFKIDIVVLAFOK	180
QY	207	ASSIVYKKEGEQVEFSPLAFTVEKLTGSGELMWAERASSKSWITFDLNKKEVSXRV	266
DB	181	ASSIVYKKEGEQVEFSPLAFTVEKLTGSGELMWAERASSKSWITFDLNKKEVSXRV	240
QY	267	TQDPKLQWKKLPLHLTLPOALPOYAGSGNLTLEAKTGKLEHVNLVVMRATQLOKNTL	326
DB	241	TQDPKLQWKKLPLHLTLPOALPOYAGSGNLTLEAKTGKLEHVNLVVMRATQLOKNTL	300
QY	327	TCEVWGPTSPKLMLSLKLENKAQVSKREKPVWVLNPEAGMOCILSDSGVLLSNIKV	386
DB	301	TCEVWGPTSPKLMLSLKLENKAQVSKREKPVWVLNPEAGMOCILSDSGVLLSNIKV	360
QY	387	LPTWSTPV 394	
DB	361	LPTWSTPV 368	

Search completed: August 3, 2004, 13:08:29
Job time : 74.926 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using bw model

Run on: August 3, 2004, 12:51:44 ; Search time 44.4955 seconds
(without alignment)
4594.975 Million cell updates/sec

Title: SEQ7
Perfect score: 3414
Sequence: 1 MNRGVPRHLLVLQALLP.....DETCAGQDELGLMTTDP 648

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 125 summaries

Database :

1: SPREMBL_25:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_organelle:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	ID	Description
1	1705	49.9	397	6	009261
2	1701	49.8	397	6	009260
3	1700	49.8	397	6	095N69
4	1696	49.7	397	6	009259
5	1696	49.7	397	6	009259
6	1689	49.5	397	6	009262
7	1688	49.4	397	6	009262
8	1601.5	46.9	457	6	08H2T8
9	1575.5	46.1	457	6	08H2T7
10	1257	36.8	509	4	08NF17
11	1221	35.8	470	4	07Z5M1
12	1214	35.6	679	4	096P08
13	1211.5	35.5	469	4	07Z7P5
14	1208.5	35.4	482	4	07Z351
15	1187	34.8	455	6	09X578
16	1163	34.1	354	4	086T72

17	1160	34.0	521	4	08N4Y9
18	1130	33.1	473	4	08TC63
19	1087	31.8	474	6	P79355
20	985.5	28.9	457	11	061396
21	958	26.1	433	11	055054
22	889	26.0	337	6	095M34
23	854	25.0	470	11	07TMT1
24	841.5	24.6	437	11	09RIA4
25	841.5	24.6	469	11	08R3V9
26	835	24.5	463	11	099LC4
27	819	24.0	468	11	099J31
28	805.5	23.6	473	11	09DBL4
29	797.5	23.4	473	11	099125
30	795.5	23.3	473	11	091205
31	794.5	23.3	474	11	08R3H6
32	416.5	12.2	613	4	096E10
33	406.5	11.9	613	4	08WUX1
34	400	11.7	86	6	077596
35	400	11.7	86	6	077597
36	400	11.7	614	11	07TMT6
37	397	11.6	86	6	077594
38	397	11.6	86	6	077599
39	395	11.6	86	6	077595
40	392	11.5	86	6	077598
41	391.5	11.5	597	4	096BB9
42	388	11.4	86	6	077601
43	388	11.4	614	4	096GN6
44	383	11.2	86	6	077600
45	382	11.2	618	4	096AA6
46	380.5	11.1	613	11	08VCX7
47	377.5	11.1	597	4	09BU10
48	375.5	11.0	597	4	09BOB8
49	372.5	10.9	588	4	08WUX4
50	357	10.5	71	4	013969
51	356.5	10.4	375	4	086T11
52	326.5	9.6	684	13	090544
53	324.5	9.5	587	13	07TOR1
54	313.5	9.2	478	4	07Z379
55	305.5	8.9	99	6	029027
56	302	8.8	416	4	09NPE6
57	300.5	8.8	494	4	096K68
58	298	8.7	496	4	096KX8
59	297	8.7	492	4	07Z374
60	296	8.7	499	4	08NSK4
61	294	8.6	493	4	08NCL6
62	289	8.5	500	4	09BRV0
63	288.5	8.5	496	4	096DK0
64	288	8.4	488	11	091WR1
65	282.5	8.3	384	4	09UP60
66	282.5	8.3	487	11	080217
67	282.5	8.3	497	4	08W124
68	280.5	8.2	99	6	029028
69	278.5	8.2	489	11	08VCX4
70	276	8.1	481	11	091WT1
71	276	8.1	482	11	08K172
72	274.5	8.0	487	11	099KX4
73	274	8.0	484	11	08VEA0
74	272	8.0	481	11	091WT3
75	272	8.0	482	11	091X92
76	271.5	8.0	479	11	091WPS
77	271.5	8.0	482	13	090WBS
78	271.5	8.0	484	11	099LAE
79	271	7.9	120	11	088650
80	270.5	7.9	486	11	091Z07
81	270	7.9	488	11	08OCF2
82	269.5	7.9	426	11	09DCD9
83	268.5	7.9	480	11	091XEL
84	259.5	7.6	481	11	08VCV5
85	255	7.5	479	11	07TMT4
86	254.5	7.5	480	11	08K0Z4
87	253.5	7.4	479	11	099M22
88	240	7.0	487	13	096WV7
89	237	6.9	90	11	P70443

08N4Y9	homo sapien
08TC63	homo sapien
P79355	Felis silve
061396	mus musculus
055054	mus musculus
095M34	equus caball
07TMT1	mus musculus
09RIA4	mus musculus
08R3V9	mus musculus
099LC4	mus musculus
099J31	mus musculus
09DBL4	mus musculus
099125	mus musculus
091205	mus musculus
08R3H6	mus musculus
096E10	homo sapien
08WUX1	homo sapien
077596	mandrillus
077597	mandrillus
07TMT6	mus musculus
077594	cercopithec
077599	cercopithec
077595	cercopithec
077598	papio sp. (
096BB9	homo sapien
077601	lophocobus
096GN6	homo sapien
077600	lophocobus
096AA6	homo sapien
08VCX7	mus musculus
09BU10	homo sapien
09BOB8	homo sapien
08WUX4	homo sapien
013969	homo sapien
086T11	homo sapien
090544	ginglymystro
07TOR1	xenopus lae
07Z379	homo sapien
029027	sus scrofa
09NPE6	homo sapien
096K68	homo sapien
096KX8	homo sapien
07Z374	homo sapien
08NSK4	homo sapien
08NCL6	homo sapien
09BRV0	homo sapien
096DK0	homo sapien
091WR1	homo sapien
09UP60	homo sapien
080217	mus musculus
08W124	homo sapien
029028	sus scrofa
08VCX4	mus musculus
091WT1	mus musculus
08K172	mus musculus
099KX4	mus musculus
08VEA0	mus musculus
091WT3	mus musculus
091X92	mus musculus
091WPS	mus musculus
090WBS	mus musculus
099LAE	mus musculus
088650	mus musculus
091Z07	mus musculus
08OCF2	mus musculus
09DCD9	mus musculus
091XEL	mus musculus
08VCV5	mus musculus
07TMT4	mus musculus
08K0Z4	mus musculus
099M22	mus musculus
096WV7	gallus gall
P70443	mus musculus

```

90 198 5 8 5636 4 096RW7 096RW7 homo sapien
91 186 5 4 258 13 090524 090524 ginglymosto
92 185.5 5 4 573 4 08WU38 08WU38 homo sapien
93 183.5 5 4 130 11 0908W4 0908W4 mus musculu
94 183.5 5 4 234 4 08N355 08N355 homo sapien
95 183.5 5 4 225 11 099M11 099M11 mus musculu
96 182.5 5 3 6620 4 096AA2 096AA2 homo sapien
97 179.5 5 3 2693 5 08ISF3 08ISF3 caenorhabd
98 179.5 5 3 2708 5 08ISF4 08ISF4 caenorhabd
99 179.5 5 3 2780 5 08MNS0 08MNS0 caenorhabd
100 179.5 5 3 18519 5 08ISF6 08ISF6 caenorhabd
101 179.5 5 3 18519 5 08ISF7 08ISF7 caenorhabd
102 179.5 5 2 18534 5 08ISF7 08ISF7 caenorhabd
103 178 5 2 17352 5 095YM2 095YM2 canis fami
104 178 5 2 17352 5 095YM2 095YM2 procambus
105 176.5 5 2 233 4 096T69 096T69 homo sapien
106 176.5 5 2 259 13 090530 090530 ginglymosto
107 175.5 5 1 1323 13 008476 008476 gallus gall
108 174.5 5 1 739 4 08N5F4 08N5F4 gallus gall
109 174.5 5 1 739 11 063669 063669 rattus norv
110 174.5 5 1 2673 4 096SC3 096SC3 homo sapien
111 174 5 1 236 4 096S61 096S61 homo sapien
112 174 5 1 8081 5 072120 072120 caenorhabd
113 172.5 5 1 234 4 0722U7 0722U7 homo sapien
114 172.5 5 1 4162 13 098918 098918 gallus gall
115 172.5 5 1 7962 4 010465 010465 homo sapien
116 172.5 5 1 34350 4 08WZ42 08WZ42 homo sapien
117 172 5 0 233 4 08TBC9 08TBC9 homo sapien
118 172 5 0 1031 13 090YM2 090YM2 brachydanto
119 171.5 5 0 257 13 090536 090536 ginglymosto
120 171 5 0 650 6 09GKR2 09GKR2 bos taurus
121 171 5 0 739 6 09GKR3 09GKR3 bos taurus
122 170 5 0 237 13 090545 090545 ginglymosto
123 170 5 0 238 7 09MXA2 09MXA2 aulonocara
124 170 5 0 379 11 08BLX5 08BLX5 mus musculu
125 170 5 0 397 11 08BFX8 08BFX8 mus musculu

```

ALIGNMENTS

```

RESULT 1
009261 PRELIMINARY; PRT; 397 AA.
AC 009261;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE CD4 (Fragment) .
GN CD4.
OS Cercopithecus sabaues.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
OC Cercopitheciinae; Cercopithecus.
OX NCBI_TaxId=60711;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98017879; PubMed=9379478;
RA Fomsgaard A., Muller-Trutwin M.C., Diop O., Hansen J., Mathiot C.,
RT Corbet S., Barre-Sinoussi F., Allan J.S.;
RT "Relation between phylogeny of African green monkey CD4 genes and
RT their respective simian immunodeficiency virus genes.";
RL J. Med. Primatol. 26:120-128(1997).
DR EMBL; AF001225; AAB60872.1; -.
DR HSSP; P01730; IMIO.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0006955; P:immune response; IEA.
DR InterPro; IPR000973; CD4_TCM6.
DR InterPro; IPR007110; IG_Like.
DR Pfam; PF00047; IG_2.
DR PRINTS; PR00692; CD4TCANTIGEN.
DR SMART; SM00406; IG; 1.
DR SMART; SM00406; IG; 1.

```

```

DR PROSITE; PSS0835; IG_Like; 1.
FT NON_TER 1
FT NON_TER 397
SQ SEQUENCE 397 AA; 43945 MW; 80C963B92A868CD3 CRC64;
Query Match
Best local similarity 89.9%; Pred. No. 1.4e-122;
Matches 330; Conservative 15; Mismatches 22; Indels 0; Gaps 0;
QY 28 VILGKGDVLELTCTASQKSIQFIWKNKSNQIKIIGNQSFLTKGPKLNDRAISRSLW 87
1 VILGKGDVLELTCTASQKSIQFIWKNKSNQIKIIGNQSFLTKGPKLNDRAISRSLW 60
QY 88 DQGNFLIKNLKIDSDPTICEVEDDQREVLTGFLRANSDTHLQGSITLTRESPP 147
61 DQGNFLIKNLKIDSDPTICEVEDDQREVLTGFLRANSDTHLQGSITLTRESPP 120
QY 148 GSSPSVQCRSPRGNKIOGGKTLVSQLELDSDGTCTVLQNKVPEKIDIVLAFQKA 207
121 GSSPSVQCRSPRGNKIOGGKTLVSQLELDSDGTCTVLQNKVPEKIDIVLAFQKA 180
QY 208 SSIIVKKEGEQVEFSPPLAFVETKLTGSGELMWQBRASSSGWITFDLKNKEVYKRYT 267
181 SSIIVKKEGEQVEFSPPLAFVETKLTGSGELMWQBRASSSGWITFDLKNKEVYKRYT 240
QY 268 ODPKLOMGKPLHLITLPALPOYAGSGNLTLALAKGKHLQENVLYVMRATOLQKNLT 327
241 ODPKLOMGKPLHLITLPALPOYAGSGNLTLALAKGKHLQENVLYVMRATOLQKNLT 300
QY 328 CEWGPSTPKMLSLKLENKAKVSRKRPVWLVNPEAGMOCLLSDSGOVLLSNIKVL 387
301 CEWGPSTPKMLSLKLENKAKVSRKRPVWLVNPEAGMOCLLSDSGOVLLSNIKVL 360
QY 388 PTWSTPV 394
361 PTWSTPV 367
DB

```

RESULT 2

```

009260 PRELIMINARY; PRT; 397 AA.
AC 009260;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE CD4 (Fragment) .
GN CD4.
OS Cercopithecus sabaues.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
OC Cercopitheciinae; Cercopithecus.
OX NCBI_TaxId=60711;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98017879; PubMed=9379478;
RA Fomsgaard A., Muller-Trutwin M.C., Diop O., Hansen J., Mathiot C.,
RT Corbet S., Barre-Sinoussi F., Allan J.S.;
RT "Relation between phylogeny of African green monkey CD4 genes and
RT their respective simian immunodeficiency virus genes.";
RL J. Med. Primatol. 26:120-128(1997).
DR EMBL; AF001224; AAB60871.1; -.
DR HSSP; P01730; IMIO.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0006955; P:immune response; IEA.
DR InterPro; IPR000973; CD4_TCM6.
DR InterPro; IPR007110; IG_Like.
DR InterPro; IPR003596; IG_V.
DR Pfam; PF00047; IG_2.
DR PRINTS; PR00692; CD4TCANTIGEN.
DR SMART; SM00406; IG; 1.
DR PROSITE; PSS0835; IG_Like; 1.
FT NON_TER 1
FT NON_TER 397

```

SQ SEQUENCE 397 AA; 43882 MW; 4788277E92EEB9 CRC64;
 Query Match 49.8%; Score 1701; DB 6; Length 397;
 Best Local Similarity 89.6%; Pred. No. 2.8e-122;
 Matches 329; Conservative 16; Mismatches 22; Indels 0; Gaps 0;

QY 28 VVLGKKGDVVELTCTASQKKSIOFHWNKSNQIKILGNQGSFLTGPSEKLNDRADSRSLW 87
 1 VVLGKKGDVVELTCTASQKKSIOFHWNKSNQIKILGNQGSFLTGPSEKLNDRADSRSLW 60
 DB 88 DQGNFLIKNLKIIEBDDTYICEVEDOKEVQLVFLTANSTHLLQGOSLTLTLESPP 147
 61 DQCFSMIKNLKIIEBDDTYICEVENKEEVELVFLTANSTHLLQGOSLTLTLESPP 120
 DB 148 GSSPSVQCSPPRGKNIQGGKTLISVQLBLQDSGTWCTVLOQKVEFKIDIVLAFOKA 207
 121 GSSPSVQCSPPRGKNIQGGKTLISVQLBLQDSGTWCTVLOQKVEFKIDIVLAFOKA 180
 DB 208 SSTVYKKEGQVFEFPLAFTVEKLTGSGELMWQAEARASSKSWITFDLNKKEVSVKRYT 267
 181 SSTVYKKEGQVFEFPLAFTVEKLTGSGELMWQAEARASSKSWITFDLNKKEVSVKRYT 240
 QY 268 QDPKLGKGLPPLTLPLPOLPYAGSGNLTLLAEKTKLHOEVNLVWMRATQLOKNT 327
 241 QDPKLGKGLPPLTLPLPOLPYAGSGNLTLLAEKTKLHOEVNLVWMRATQLOKNT 300
 DB 328 CEVWGPTSPKLMSTLKENKAVSKREKPVWVNLNPEAGMOCCLSDSGVLLSNIKYL 387
 301 CEVWGPTSPKLMSTLKENKAVSKREKPVWVNLNPEAGMOCCLSDSGVLLSNIKYL 360
 QY 388 PTWSTPV 394
 DB 361 PTWSTPV 367

RESULT 3
 ID 095NE9 PRELIMINARY; PRT; 397 AA.
 AC 095NE9;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE CD4 (Fragment).
 GN CD4.
 OS Cercopithecus pygerythrus.
 OC Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecinae; Cercopithecus.
 NCBI_TaxID=60710;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98017879; PubMed=9379478;
 RA Fomsgaard A., Muller-Trutwin M.C., Diop O., Hansen J., Mathiot C.,
 Corbet S., Barre-Sinoussi F., Allan J.S.;
 RT "Relation between phylogeny of African green monkey CD4 genes and
 their respective simian immunodeficiency virus genes.";
 RL J. Med. Primatol. 26:120-128(1997).
 DR EMBL; AF001227; AAB60874.1; -.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0006955; P:immune response; IEA.
 DR InterPro; IPR000973; CD4_TcAg.
 DR InterPro; IPR003596; Ig_V.
 DR Pfam; PF00047; Ig_2.
 DR SMART; SM00406; IGV; 1.
 DR PRINTS; PR00692; CD4TCANTIGEN.
 DR SMART; SM00406; IGV; 1.
 DR PROSITE; PSS0835; IG_LIKE; 1.
 FT NON_TER 1
 FT NON_TER 1
 FT SEQUENCE 397 AA; 43946 MW; 21C3530882ABFC0 CRC64;
 SQ SEQUENCE 397 AA; 43946 MW; 21C3530882ABFC0 CRC64;
 Query Match 49.8%; Score 1700; DB 6; Length 397;
 Best Local Similarity 89.6%; Pred. No. 3.3e-122;

Matches 329; Conservative 15; Mismatches 23; Indels 0; Gaps 0;

QY 28 VVLGKKGDVVELTCTASQKKSIOFHWNKSNQIKILGNQGSFLTGPSEKLNDRADSRSLW 87
 1 VVLGKKGDVVELTCTASQKKSIOFHWNKSNQIKILGNQGSFLTGPSEKLNDRADSRSLW 60
 DB 88 DQGNFLIKNLKIIEBDDTYICEVEDOKEVQLVFLTANSTHLLQGOSLTLTLESPP 147
 61 DQCFSMIKNLKIIEBDDTYICEVENKEEVELVFLTANSTHLLQGOSLTLTLESPP 120
 DB 148 GSSPSVQCSPPRGKNIQGGKTLISVQLBLQDSGTWCTVLOQKVEFKIDIVLAFOKA 207
 121 GSSPSVQCSPPRGKNIQGGKTLISVQLBLQDSGTWCTVLOQKVEFKIDIVLAFOKA 180
 DB 208 SSTVYKKEGQVFEFPLAFTVEKLTGSGELMWQAEARASSKSWITFDLNKKEVSVKRYT 267
 181 SSTVYKKEGQVFEFPLAFTVEKLTGSGELMWQAEARASSKSWITFDLNKKEVSVKRYT 240
 QY 268 QDPKLGKGLPPLTLPLPOLPYAGSGNLTLLAEKTKLHOEVNLVWMRATQLOKNT 327
 241 QDPKLGKGLPPLTLPLPOLPYAGSGNLTLLAEKTKLHOEVNLVWMRATQLOKNT 300
 DB 328 CEVWGPTSPKLMSTLKENKAVSKREKPVWVNLNPEAGMOCCLSDSGVLLSNIKYL 387
 301 CEVWGPTSPKLMSTLKENKAVSKREKPVWVNLNPEAGMOCCLSDSGVLLSNIKYL 360
 QY 388 PTWSTPV 394
 DB 361 PTWSTPV 367

RESULT 4
 ID 009259 PRELIMINARY; PRT; 397 AA.
 AC 009259;
 DT 01-JUL-1997 (TrEMBLrel. 04, Created)
 DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE CD4 (Fragment).
 GN CD4.
 OS Cercopithecus aethiops.
 OC Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecinae; Cercopithecus.
 NCBI_TaxID=60711;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98017879; PubMed=9379478;
 RA Fomsgaard A., Muller-Trutwin M.C., Diop O., Hansen J., Mathiot C.,
 Corbet S., Barre-Sinoussi F., Allan J.S.;
 RT "Relation between phylogeny of African green monkey CD4 genes and
 their respective simian immunodeficiency virus genes.";
 RL J. Med. Primatol. 26:120-128(1997).
 DR EMBL; AF001223; AAB60870.1; -.
 DR HSSP; P01730; IMIO.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0006955; P:immune response; IEA.
 DR InterPro; IPR000973; CD4_TcAg.
 DR InterPro; IPR001110; Ig_Like.
 DR InterPro; IPR003596; Ig_V.
 DR Pfam; PF00047; Ig_2.
 DR PRINTS; PR00692; CD4TCANTIGEN.
 DR SMART; SM00406; IGV; 1.
 DR PROSITE; PSS0835; IG_LIKE; 1.
 FT NON_TER 1
 FT NON_TER 1
 FT SEQUENCE 397 AA; 43881 MW; 7CE39AD0F8506C81 CRC64;
 SQ SEQUENCE 397 AA; 43881 MW; 7CE39AD0F8506C81 CRC64;
 Query Match 49.7%; Score 1696; DB 6; Length 397;
 Best Local Similarity 89.4%; Pred. No. 6.7e-122;
 Matches 328; Conservative 17; Mismatches 22; Indels 0; Gaps 0;

QY 28 VVLGKKGDVVELTCTASQKKSIOFHWNKSNQIKILGNQGSFLTGPSEKLNDRADSRSLW 87

```

Db      1 VVLGKGGDVELTLCNASQKTTQFHHKNSNQIKILKQSGFLTKSSKLRDRIDSRKSLW 60
Qy      88 DQGNFPLIIKNIKIEDSDTYICEVEDQKEEVLVFGLTANSPTHLLQGOSLTLTLESP 147
        61 DQGCFSMIITKNIKIEDSEITYICEVENKEEVELVGLTANSPTHLLQGOSLTLTLESP 120
Db      148 GSSPSVQCRSPRGKNIQGGKTLVSQLELODSGTWCTVLQONKVEFKIDIVLAFOKA 207
        121 GSSPSVKCRSPRGKNIQGGRTLSVPQLERODSGTWCTVSNQONVTEFKIDIVLAFOKA 180
Qy      208 SSTVYKKEGEQVEFSFPLAFVTEKLTGSGELMWQAEARASSSKSWITFDLKNKESVKRT 267
        181 SSTVYKKEGEQVEFSFPLAFVTEKLTGSGELMWQAEARASSSKSWITFDLKNKESVKQVT 240
Db      268 QDPKLOMGKKLPLHLTLPOALPOYAGSGNLTALBAKTKLHQEVNLVVMRATQLOKNTL 327
        241 QDPKLOMGKKLPLHLTLPOALPOYAGSGNLTALBAKTKLHQEVNLVVMRATQLOFENLT 300
Qy      328 CEVWGPTSPKMLSLKLENKEAKVSKREKPVWVLNPEAGMOCLLSDSGQVLLSNIKVL 387
        301 CEVWGPTSPKMLSLKLENKAATVSKAKAVWVLNPEAGMOCLLSDSGQVLLSNIKVL 360
Db      388 PTWSTPV 394
        361 PTWPTPV 367

```

RESULT 5

```

009263 ID 009263 PRELIMINARY; PRT; 397 AA.

```

```

AC 009263; 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE CD4 (Fragment).
GN GN
OS Cercopithecus tantalus.
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC Cercopithecinae; Cercopithecus.
OX NCBI_TaxID=60712;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98017879; PubMed=9379478;
RA Fomsgaard A., Muller-Trutwin M.C., Diop O., Hansen J., Mathiot C.,
RA Corbet S., Barre-Sinoussi F., Allan J.S.;
RT "Relation between phylogeny of African green monkey CD4 genes and
RT their respective simian immunodeficiency virus genes.";
RL J. Med. Primatol. 26:120-128(1997).
DR EMBL; AF001221; AAB60869.1; -.
DR HSSP; P01730; IMIQ.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0006955; P:immune response; IEA.
DR InterPro; IPR000973; CD4_TcAg.
DR InterPro; IPR007110; Ig_Like.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig_2.
DR PRINTS; PR00692; CD4TCANTIGEN.
DR SMART; SM00406; IGV_1.
DR PROSITE; PSS0835; IG_LIKE; 1.
FT NON_TER 1
FT NON_TER 397
SQ SEQUENCE 397 AA; 43994 MW; A3CD031535A51524 CRC64;

```

Query Match 49.7%; Score 1696; DB 6; Length 397;

Best Local Similarity 89.4%; Pred. No. 6.7e-122; Indels 0; Gaps 0;

Matches 328; Conservative 15; Mismatches 24; Indels 0; Gaps 0;

```

Qy      88 DQGNFPLIIKNIKIEDSDTYICEVEDQKEEVLVFGLTANSPTHLLQGOSLTLTLESP 147
        61 DQGCFSMIITKNIKIEDSEITYICEVENKEEVELVGLTANSPTHLLQGOSLTLTLESP 120
Db      148 GSSPSVQCRSPRGKNIQGGKTLVSQLELODSGTWCTVLQONKVEFKIDIVLAFOKA 207
        121 GSSPSVKCRSPRGKNIQGGRTLSVPQLERODSGTWCTVSNQONVTEFKIDIVLAFOKA 180
Qy      208 SSTVYKKEGEQVEFSFPLAFVTEKLTGSGELMWQAEARASSSKSWITFDLKNKESVKRT 267
        181 SSTVYKKEGEQVEFSFPLAFVTEKLTGSGELMWQAEARASSSKSWITFDLKNKESVKQVT 240
Db      268 QDPKLOMGKKLPLHLTLPOALPOYAGSGNLTALBAKTKLHQEVNLVVMRATQLOKNTL 327
        241 QDPKLOMGKKLPLHLTLPOALPOYAGSGNLTALBAKTKLHQEVNLVVMRATQLOFENLT 300
Qy      328 CEVWGPTSPKMLSLKLENKEAKVSKREKPVWVLNPEAGMOCLLSDSGQVLLSNIKVL 387
        301 CEVWGPTSPKMLSLKLENKAATVSKAKAVWVLNPEAGMOCLLSDSGQVLLSNIKVL 360
Db      388 PTWSTPV 394
        361 PTWPTPV 367

```

RESULT 6

```

009262 ID 009262 PRELIMINARY; PRT; 397 AA.

```

```

AC 009262; 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE CD4 (Fragment).
GN GN
OS Cercopithecus tantalus.
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC Cercopithecinae; Cercopithecus.
OX NCBI_TaxID=60712;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98017879; PubMed=9379478;
RA Fomsgaard A., Muller-Trutwin M.C., Diop O., Hansen J., Mathiot C.,
RA Corbet S., Barre-Sinoussi F., Allan J.S.;
RT "Relation between phylogeny of African green monkey CD4 genes and
RT their respective simian immunodeficiency virus genes.";
RL J. Med. Primatol. 26:120-128(1997).
DR EMBL; AF001221; AAB60868.1; -.
DR HSSP; P01730; IMIQ.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0006955; P:immune response; IEA.
DR InterPro; IPR000973; CD4_TcAg.
DR InterPro; IPR007110; Ig_Like.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig_2.
DR PRINTS; PR00692; CD4TCANTIGEN.
DR SMART; SM00406; IGV_1.
DR PROSITE; PSS0835; IG_LIKE; 1.
FT NON_TER 1
FT NON_TER 397
SQ SEQUENCE 397 AA; 43954 MW; CF7F2F5D82335B0D CRC64;

```

Query Match 49.5%; Score 1689; DB 6; Length 397;

Best Local Similarity 89.4%; Pred. No. 2.3e-121; Indels 0; Gaps 0;

Matches 328; Conservative 15; Mismatches 24; Indels 0; Gaps 0;

28 VVLGKGGDVELTLCNASQKSIQFHHKNSNQIKILKQSGFLTKGSKLRDRADSRRLW 87

1 VVLGKGGDVELTLCNASQNTTQFHHKNSNQIKILKQSGFLTKGSKLRDRIDSRKSLW 60

88 DQGNFPLIIKNIKIEDSDTYICEVEDQKEEVLVFGLTANSPTHLLQGOSLTLTLESP 147

61 DQGCFSMIITKNIKIEDSEITYICEVENKEEVELVGLTANSPTHLLQGOSLTLTLESP 120

```

Qy 148 GSSPSVOCRRSPRGKNIQGGKTLVSQLELSDSGTWTCTVLOQKKEFKIDIVLAFOKA 207
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 GSSPSVOCRRSPRGKNIQGGKTLVSQLELSDSGTWTCTVLOQKKEFKIDIVLAFOKA 180
Qy 208 SSIYKKEGEQVEFSFPLAFTVEKLTGSGELMWOAERASSSSKSWITFDLKNKEVSVRVT 267
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 SSIYKKEGEQVEFSFPLAFTVEKLTGSGELMWOAERASSSSKSWITFDLKNKEVSVRVT 240
Qy 268 QDPKLQMGKKLPHLTLTLPALPOYAGSGNLTALAEATGKLGHOEVLVWVRATQLOKRLT 327
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 241 QDPKLQMGKKLPHLTLTLPALPOYAGSGNLTALAEATGKLGHOEVLVWVRATQLOKRLT 300
Qy 328 CEVWGPTSPKLMLSLKLENKAYSKREKPYVWLNPRAGMWQCILSDSGVLESNIKVL 387
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 301 CEVWGPTSPKLMLSLKLENKAYSKREKPYVWLNPRAGMWQCILSDSGVLESNIKVL 360
Qy 388 PTWSTPV 394
    |||||
Db 361 PTWSTPV 367

```

RESULT 7

```

ID 002805 PRELIMINARY; PRT; 397 AA.
AC 002805; 077593;
DT 01-JUL-1997 (Tremblrel. 04, Created)
DT 01-JUL-1997 (Tremblrel. 04, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/LEU-3)
    (Fragment).
GN CD4.
OS Cercopithecus aethiops (Green monkey) (Griwet).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Cercopithecus.
OX NCBI_TaxID=9534;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Peripheral blood;
RA MEDLINE=98017879; PubMed=9379478;
RA Fomsgard A., Møller-Trutwin M.C., Diop O., Hansen J., Mathiot C.,
RA Corbet S., Barre-Sinoussi F., Allan J.S.;
RT "Relation between phylogeny of African green monkey CD4 genes and
RT their respective simian immunodeficiency virus genes.";
RL J. Med. Primatol. 26:120-128(1997).
RN [2]
RP SEQUENCE OF 80-165 FROM N.A.
RX MEDLINE=98320644; PubMed=9656488;
RA Harrie E.E., Diaciell T.R.;
RT "Nuclear gene trees and the phylogenetic relationships of the
RT mangabeys (Primates: Papionini).";
RL Mol. Biol. Evol. 15:892-900(1998).
CC -1- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL
CC RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY
CC SIMILARITY).
CC -1- SUBUNIT: ASSOCIATES WITH P56-LCK (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC ONE V-LIKE AND 3 C2-LIKE DOMAINS.
DR EMBL: AF001226; AAB60873.1; -
DR EMBL: AF057380; AAC5124.1; -
DR HSP; P01730; IWIQ.
DR GO: GO:0016021; C:Integral to membrane; IEA.
DR GO: GO:0006955; P:Immune response; IEA.
DR InterPro: IPR000973; CD4_TGAG.
DR InterPro: IPR007110; IG-Like.
DR InterPro: IPR003596; IG_V.
DR Pfam; PF00047; 19; 2.
DR PRINTS; PR00692; CD4TCANTIGEN.
DR SMART; SM00406; IGV.1.
KM Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
    Lipoprotein; Palmitate; Repeat.

```

```

FT NON TER 1 1
FT DOMAIN <1 370
FT TRANSMEM 371 391
FT DOMAIN 392
FT DOMAIN 392 >397
FT DOMAIN <1 98
FT DOMAIN 99 176
FT DOMAIN 177 290
FT DOMAIN 291 347
FT CARBOHYD 15 15
FT CARBOHYD 30 30
FT CARBOHYD 269 269
FT CARBOHYD 298 298
FT DISULFID 14 82
FT DISULFID 128 157
FT DISULFID 301 343
FT LIPID 392 392
FT LIPID 395 395
FT LIPID 397
FT NON TER 397
SQ SEQUENCE 397 AA; 43980 MW; F74C42E2B196155 CRC64;

```

Query Match 49.4%; Score 1688; DB 6; Length 397;
 Best Local Similarity 89.1%; Pred. No. 2.8e-121;
 Matches 327; Conservative 15; Mismatches 25; Indels 0; Gaps 0;

```

Qy 28 VLVGKGDVTELTCTASQKKSIOFHWKNSQIKLGQSGFLTKGPKLNDRADSRSLW 87
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 VLVGKGDVTELTCTASQKKTTOFHWKNSQIKLGQSGFLTKGSKLNDRIDSRSLW 60
Qy 88 DQGNFPLIINKLIEDSDIYICEVEDQKEEVQLLVFGLTANSPTHLQGGSLTLTLESP 147
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 DQGNFPLIINKLIEDSEYICEVENKEVEVLVFGLTANSPTHLQGGSLTLTLESP 120
Qy 148 GSSPSVOCRRSPRGKNIQGGKTLVSQLELSDSGTWTCTVLOQKKEFKIDIVLAFOKA 207
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 GSSPSVOCRRSPRGKNIQGGKTLVSQLELSDSGTWTCTVLOQKKEFKIDIVLAFOKA 180
Qy 208 SSIYKKEGEQVEFSFPLAFTVEKLTGSGELMWOAERASSSSKSWITFDLKNKEVSVRVT 267
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 SSIYKKEGEQVEFSFPLAFTVEKLTGSGELMWOAERASSSSKSWITFDLKNKEVSVRVT 240
Qy 268 QDPKLQMGKKLPHLTLTLPALPOYAGSGNLTALAEATGKLGHOEVLVWVRATQLOKRLT 327
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 241 QDPKLQMGKKLPHLTLTLPALPOYAGSGNLTALAEATGKLGHOEVLVWVRATQLOKRLT 300
Qy 328 CEVWGPTSPKLMLSLKLENKAYSKREKPYVWLNPRAGMWQCILSDSGVLESNIKVL 387
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 301 CEVWGPTSPKLMLSLKLENKAYSKREKPYVWLNPRAGMWQCILSDSGVLESNIKVL 360
Qy 388 PTWSTPV 394
    |||||
Db 361 PTWSTPV 367

```

RESULT 8

```

ID 08H278 PRELIMINARY; PRT; 457 AA.
AC 08H278;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Lymphocyte antigen CD4.
OS Callithrix jacchus (Common marmoset).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae; Callitrich.
OX NCBI_TaxID=9483;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22174698; PubMed=12186836;
RA Labonte J.A., Babcock G.J., Patel T., Sodroski J.;
RT "Blockade of HIV-1 Infection of New World Monkey Cells Occurs
RT Primarily at the Stage of Virus Entry.";
RL J. Exp. Med. 196:431-445(2002).
DR EMBL: AF452616; AAN14532.1; -

```

PLAIN; FF0004// 1g; 2.

Query match	36.86;	Score 1257;	DB 4;	Length 509;
Best local similarity	89.38;	Pred. No. 5.1e-88;		

Query match	36.86;	Score 125/;	DB 4;	Length 509;
Best local similarity	89.38;	Pred. No. 5.1e-88;		

	Matches	233;	Conservative	11;	Mismatches	11;	IndeIs	6;	Gaps	2
Oy	332	TVPFGP-APBPSCSKTHTC----	PBLIGPSVFLPPPKQDTLMISRTPEVTGVVDY	445						
Dd	200	TPPCPRCPBPSPKSCDTPPCPCRCPA	BELIIGGVSVELFPFKPDOTLMISTRTEVCVVDD	259						
Oy	446	SHEDEVEFNWVVDDEVHNAKTAKR	EONYSIVTSVLTYLHDMDLNKGXKKCVSNK	505						
Dd	260	SHEDEVGPKMVVDGEVNAAKTKREB	OFNSTRFVSVSLTVLHODMNLNGEXKKCVSNK	319						
Oy	506	ALPALIEKTISSKACQPREFPVYTTP	SRDELTKIQNVSLTCLVKFYPSDIAVENESNQ	565						
Dd	320	ALPALIEKIISTKQQOPREPQYYTL	LPRESERTIKQNQLSTCLVKGFYSIDIAVENESSQG	379						
Oy	566	PENNYKTTTPVIDSGSFVLSKLTVDS	KRWOGNSVFCSVMHEALHNHYTOKSLSISP	625						
Dd	380	PENNNTPTPMIDSGSEFFLYSKLTVD	XSRMOGGNI FSCSVMHBALNHFRFTQSLSLBE	439						
Oy	626	LQLDCTCAADGDGLDLMTT	646							
Dd	440	LQLESCAABOQDGELDGLMTT	460							
RESULT 11										
ID	07ZSW1	PRELIMINARY:	PRT:	470 AA.						
AC	07ZSW1,									
DT	01-OCT--2003 (TREMBLrel. 25,	Created)								
DR	01-OCT--2003 (TREMBLrel. 25,	Last sequence update)								
DE	01-OCT--2003 (TREMBLrel. 25,	Last annotation update)								
DI	Hypothetical protein.									
OS	Homo sapiens (Human).									
OC	Eukaryota; Metazoa; Chordata;	Cranialata; Vertebrata; Euteleostomi;								
CC	Mammalia; Euteria; Primates;	Catarrhini; Homidae; Homo.								
OX	NCB1_Taxid=9606;	[1]								
RN	SEQUENCE FROM N.A.									
RP	TISSUE=Spleen.									
RC	MEDLINE=22388257; PubMed=12477932;									
RX	Strausberg R.L., Feingold E.A., Groube L.H., Derge J.G.,									
RA	Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,									
RB	Alechski S.F., Zeeberg B., Bucoek K.H., Scheffer C.F., Bhat N.K.,									
RD	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hasel F.,									
RF	Diachenko L., Marusina K., Farmer A.C., Rubin G.W., Hong L.,									
RG	Steggleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,									
RH	Brownstein M.J., Uediri T.B., Toshyluki S., Carlincl P., Prange C.,									
RI	Rana S.S., Logueano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,									
RJ	Bosak S.A., McEwen P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,									
RK	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huliy S.W.,									
RL	Villalon D.K., Muzy D.M., Sodergren E.U., Lu X., Gibbs R.A.,									
RM	Fahney J., Helton E., Ketteman M., Madan A., Rodriguez S., Sanchez A.,									
RN	Whiting R.W., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,									
RO	Blakeley M.W., Touchman J.W., Green E.D., Dickson M.C.,									
RR	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,									
RS	Krzywinski M.T., Skalska A., Smalins D.E., Schnorch A., Schein J.E.,									
RU	Jones S.J., Maria M.A.;									
RV	"Generation and initial analysis of more than 15,000 full-length human									
WT	cDNA sequences."									
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).									
RN	[2]									
RP	SEQUENCE FROM N.A.									
RC	TISSUE=Spleen;									
RD	Strausberg R.;									
RE	Submitted (JUN-2003) to the EMBL/genbank/dbj databases.									
RL	EMBL: BC053984; AAC53984.1; --									
RK	Hypothetical protein.									
SO	SEQUENCE 470 aa; 51204 MW; 778CF34521483E1A CRC64;									
Query Match	35.8%; Score 1221; DB 4; Length 470;									
Best Local Similarity	46.4%; Pred. No. 2,6e-85;									
Matches 284; Conservative 36; Mismatches 104; Indels 188; Gaps 15;										

Db	30	LVQPGSGSLRLSCVASGFTLANNYDMHWROGI	GKLEWVNSKIGTNA	DRRYAGSVNGRFTIS	89				
Qy	79	RADSRRLSMDGNPELLII	KNLIKIEDSDTYICEVDOKEEYVLLAFGLTANS	DTILLQOS	138				
Db	90	RENAKDSLYLOMN-----	SLRVGDAVAVYC-----	ARGAGWAPL	GARDINGOC	133			
Qy	139	LTLTLESPGSSPSVQCS	PRGKNIQGGKTL	SVSOLBDSGTWCTVLONOK	KVEFKID	198			
Db	134	TMVTVSSASTKPSGVF	PLAPSSKSTSGG-TAALCCL-----			168			
Qy	159	IVLAFQASSIYVKKGEQVEF	PEFLAFVYEKLTGSGELMWOAERASSSKSWIT	TPDLN	258				
Db	169	-----	VKDYFPEPVTYS-----	WMSGALTSG-----		189			
Qy	259	KEVSIVKVTQDPKTLQMGK	PLHLYTLQALPOYVAGS	GNLTALAEAKTGKJLH	OEYNLVMR	318			
Db	190	-----	VH-TFPAVL-QSSGLX	SLSSVTVTPSSLSGTQYI----		222			
Qy	319	ATOLQKULTCEVWGPTSP	KMLSLKENKEAKVSKREKPV	WVNLNPEAGMOC	LLSDSGOV	378			
Db	223	-----	CVN-----	NHKRSNTKVDKV-----		238			
Qy	379	LLESNIKLPFWSTPVP	PCAPBEPKSCDKTTC-----	PELLGGSVFLFP	PKPDTLMIS	433			
Db	239	-----	EPKSCDKTHTCP	PCAPBELLGGSVFLFP	PKPDTLMIS	277			
Qy	434	RTPEVTCVVDVSHEDPE	VEKFNWVVDGVEVNAATKRE	EQVNSTYVSVLTL	HDWL	493			
Db	278	RTPEVTCVVDVSHEDPE	VKFNWVVDGVEVNAATKRE	EQVNSTYVSVLTL	HDWL	337			
Qy	494	NGKEYCKVSNKALPA	PIEKTISKAKQPREPQVYTL	PSRDELTKQVSL	TLCKVGFY	553			
Db	338	NGKEYCKVSNKALPA	PIEKTISKAKQPREPQVYTL	PSRDELTKQVSL	TLCKVGFY	397			
Qy	554	SDIAVEMESNQPENNYK	TPPVLDSDGSFFLYSKL	TVDKSRMOQGNV	FSCSVMEHLN	613			
Db	398	SDIAVEMESNQPENNYK	TPPVLDSDGSFFLYSKL	TVDKSRMOQGNV	FSCSVMEHLN	457			
Qy	614	HYTOKSLSLSPG	625						
Db	458	HYTOKSLSLSPG	469						
RESULT 12									
Q96PQ8									
ID	Q96PQ8	PRELIMINARY;		PRT;	679 AA.				
AC	Q96PQ8:	01-DEC-2001 (TREMBlrel. 19, Created)							
DT	01-JUN-2003	(TREMBlrel. 24, Last sequence update)							
DT	01-OCT-2003	(TREMBlrel. 25, Last annotation update)							
DE	Factor VII active site mutant immunocjugate.								
OS	Homo sapiens (Human).								
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;								
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.								
OX	NCBI_TaxID=9606;								
RN	[1]								
RP	SEQUENCE FROM N.A.								
RX	MEDLINE=21477446;	PubMed=11593034;							
RA	Hu Z., Garen A.;								
RT	"Targeting tissue factor on tumor vascular endothelial cells and tumor cells for immunotherapy in mouse models of prostatic cancer.";								
RL	Proc. Natl. Acad. Sci. U.S.A. 98:12180-12185(2001).								
RN	[2]								
RP	SEQUENCE FROM N.A.								
RA	Hu Z., Garen A.;								
RL	Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.								
DR	EMBL; AF272774;	AAKS8686.2;							
DR	GO; GO:0005576;	C:extracellular binding.							
DR	GO; GO:0005509;	F:calcium ion binding.							
DR	GO; GO:0004263;	F:chymotrypsin activity; IEA.							
DR	GO; GO:0004295;	F:trypsin activity; IEA.							
DR	GO; GO:0006508;	P:proteolysis and peptidolysis; IEA.							

SQ SEQUENCE 482 AA; 52852 MW; EDA75F1901D1A034 CRC64;
 Query Match 35.4%; Score 1208.5; DB 4; Length 482;
 Best Local Similarity 97.4%; Pred. No. 2.5e-84;
 Matches 225; Conservative 0; Mismatches 1; Indels 5; Gaps 1;

QY 400 EPKSCDKTHTC-----PELLGSPVFLPPPKDITLMSRPEVTCVVVDVSHEDPEVKF 454
 DB 251 EPKSCDKTHTCPCPAPPELLGSPVFLPPPKDITLMSRPEVTCVVVDVSHEDPEVKF 310
 QY 455 NMVYDGVENNAKTRPREEQYNSTYRVSVLTTLHOMLNGKEYKCKVSNKALPAPIEKT 514
 DB 311 NMVYDGVENNAKTRPREEQYNSTYRVSVLTTLHOMLNGKEYKCKVSNKALPAPIEKT 370
 QY 515 ISKAGGPREFQVYTLTPSPRELTKNOVSLTCLVKGFPSPDIAVWESNGCPENNYKTPP 574
 DB 371 ISKAGGPREFQVYTLTPSPRELTKNOVSLTCLVKGFPSPDIAVWESNGCPENNYKTPP 430

QY 575 PVLDSDGSFFLYSKLTVDKSRMOQGNVFCVMEHALHNHYTOKSLSPG 625
 DB 431 PVLDSDGSFFLYSKLTVDKSRMOQGNVFCVMEHALHNHYTOKSLSPG 481

RESULT 15
 Q9XS78 PRELIMINARY; PRT; 455 AA.
 AC Q9XS78;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE T-cell surface glycoprotein CD4.
 OS Delphinapterus leucas (Beluga whale).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Odontoceti;
 OC Monodonta; Delphinapterus.
 ON NCBI_TaxID=9749;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Thymus;
 RX MEDLINE=99216435; PubMed=10199913;
 RA Romano T.A., Ridgway S.H., Felten D.L., Quaranta V.;
 RT "Molecular cloning and characterization of CD4 in an aquatic mammal,
 RT the white whale Delphinapterus leucas";
 RL Immunogenetics 49:376-383(1999).
 DR EMBL; AF071799; AAD3738.1; -;
 DR HSPB; P01730; IWIQ.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0006955; P:immune response; IEA.
 DR InterPro; IPR000973; CD4 TCAG.
 DR InterPro; IPR007110; Ig-Like.
 DR InterPro; IPR003596; Ig_V.
 DR Pfam; PF00047; Ig_3.
 DR PRINTS; PRO0692; CD4TCANTIGEN.
 DR SMART; SM00406; IGV; 1.
 DR PROSITE; PSS0835; IG_LIKE; 2.
 SQ SEQUENCE 455 AA; 50499 MW; AA532FD441AA5D1 CRC64;

Query Match 34.8%; Score 1187; DB 6; Length 455;
 Best Local Similarity 60.5%; Pred. No. 1e-82;
 Matches 251; Conservative 46; Mismatches 98; Indels 20; Gaps 5;

QY 1 MNKRVPRHLLLVQLALLPAATQGNVVLGKKDYTELTCTAOKSIOGHMKNQK 60
 DB 1 MDPTSRHLFLVQLVWLPAGTGKRVVLGKAGELPCKASQNKSLFPMNSYOTK 60
 QY 61 ILGNQGFLLTKGPKLMDRADSRSLMDQGNFLIKNLKIEDDTYICEVEDQKEVOL 120
 DB 61 ILGNHGFPMHKGANLHRSRVSKINLMDQGFPLVIDLBPBGGTIIICEVEDKIIVEL 120
 QY 121 LVPELTAUSDTHLLOQSGSLTLTPSPGSSPSVDCSRPRGNKIGQKTLVSQLELDG 180
 DB 121 QVFLTLASSDTRLILLOSLTLTLEGPSGSPSVQMKGFNGKRNKNEAKSLSPQVGLDSDG 180

QY 181 TWCTVLONOKRYEPKIDIVLAFORASSIVYKKEGQVFPPLAFTVEKLTGSGEL-W 239
 DB 181 TWCTVSOAQGTIVFNHILVLAFOVSSIVYAKKEGQMFSPPLTGTGDEL--SGELSW 238
 QY 240 WQARRASSSWITFPDKNKEVSVKRVTQDPKIQMGKPLHLTLTPOALPOVAGSGNLT 299
 DB 239 LQAKGNSSPSSWITFKLNKGKIVTGKARKDLKLRMSKALPLHLTLTPOALPOVAGSGNLT 298
 QY 300 ALEAKTGKHOENIVVMRATQLOKNTCEVMQPTSPKMLSLKLENKEKVSREKRPW 359
 DB 299 NL--YKGLYQEVNVLVVRVTKSPNSLTCEVLDPTSPRLTLISKENQSRVSDQQLVT 356
 QY 360 VLNPEAGMOCCLSDSGOVLLESNIKVLPWSTPVPCEAPDEPKSCDKTHTCPELL 414
 DB 357 VLGPPEAGMOCCLSDSGOVLLESNIKVLPWSTPVPCEAPDEPKSCDKTHTCPELL 396

RESULT 16
 Q86TT2 PRELIMINARY; PRT; 354 AA.
 AC Q86TT2;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Human full-length cDNA clone CSOD1019YF20 of placenta of Homo sapiens
 DE (Human) (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 ON NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RA Genoscope;
 RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RA Li W.B., Gruber C., Jassse J., Polayes D.;
 RT "Full-length cDNA libraries and normalization";
 RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BX48278; CAD62606.1; -;
 DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
 DR InterPro; IPR007110; Ig-Like.
 DR InterPro; IPR003597; Ig_c1.
 DR InterPro; IPR003596; Ig_MHC.
 DR Pfam; PF00047; Ig_1.
 DR SMART; SM00407; IGc1; 3.
 DR PROSITE; PSS0835; IG_LIKE; 3.
 DR PROSITE; PSS0290; IG_MHC; 2.
 KW Plasmid.
 FT NON TER
 SQ SEQUENCE 354 AA; 39125 MW; 23B80FAD2B87A92 CRC64;

Query Match 34.1%; Score 1163; DB 4; Length 354;
 Best Local Similarity 90.0%; Pred. No. 5e-81;
 Matches 216; Conservative 8; Mismatches 10; Indels 6; Gaps 2;

QY 392 TPVPCP-APPKSCDKTHTC-----PELLGSPVFLPPPKDITLMSRPEVTCVVVDV 445
 DB 114 TPVPCP-APPKSCDKTHTC-----PELLGSPVFLPPPKDITLMSRPEVTCVVVDV 173
 QY 446 SHEDPEVKFMVYDGVENNAKTRPREEQYNSTYRVSVLTTLHOMLNGKEYKCKVSNK 505
 DB 174 SHEDPEVKFMVYDGVENNAKTRPREEQYNSTYRVSVLTTLHOMLNGKEYKCKVSNK 233
 QY 506 ALPAPIEKTISKAGGPREFQVYTLTPSPRELTKNOVSLTCLVKGFPSPDIAVWESNGC 565
 DB 234 ALPAPIEKTISKAGGPREFQVYTLTPSPRELTKNOVSLTCLVKGFPSPDIAVWESNGC 293
 QY 566 PENNYKTPPVLDSDGSFFLYSKLTVDKSRMOQGNVFCVMEHALHNHYTOKSLSPG 625
 DB 294 PENNYKTPPVLDSDGSFFLYSKLTVDKSRMOQGNVFCVMEHALHNHYTOKSLSPG 353

RESULT 17

Q8N4Y9 PRELIMINARY; PRT; 521 AA.
 ID Q8N4Y9;
 AC Q8N4Y9;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 NCBI_Taxid=9606;
 RX [1]
 RN SEQUENCE FROM N.A.
 RC TISSUE=Primary B-Cells;
 RA Strausberg R.;
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC033178; AAH33178.1; -.
 DR F01; A60764; A60764.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003597; Ig_C1.
 DR InterPro; IPR003006; Ig_MHC.
 DR InterPro; IPR003596; Ig_V.
 DR Pfam; PF00047; Ig_4.
 DR SMART; SM00407; IGV; 3.
 DR SMART; SM00406; IGV; 1.
 DR PROSITE; PS00835; IG_LIKE; 4.
 DR PROSITE; PS00290; IG_MHC; 2.
 DR Hypothetical protein.
 KW SEQUENCE 521 AA; 57156 MW; 2AC7D22E7D6CAA2 CRC64;
 SQ
 Query Match 34.0%; Score 1160; DB 4; Length 521;
 Best Local Similarity 49.2%; Pred. No. 1.5e-80;
 Matches 281; Conservative 48; Mismatches 110; Indels 132; Gaps 24;
 Db
 117 EVQLVFLGKLANSDTHLQ-GQSLTLT-----LESPPGSSPS-VQC-RSP 158
 20 EVQLV-----DSGGGVPGSSLRSLCAASGFVSDHYEWRQAPGKPEWVGCRSK 73
 159 RGR-----NIQGGKT-----LSVQLELDQSGTWCT-VLQNKVEFKIDI 199
 74 AHSTTEYASVGRITILRDSKSNVHLQMSLKTDTAVYVCVADLEAGKDYVPI 133
 200 ----VLAFOKASSIYKKEGQVEFSFPLAFTVEKLTGSGELMWAERASSKSWITPD 255
 134 WGRGILVTVSSAST-----KGPV--FPLAPCSRSTSG-----GTALGCLVXD 175
 256 LKKKEYSVGRVQDPKLOMGKPLHLTLPLQALPQVAGSNTLALAEKTKL----- 308
 176 YPEPPTVS--WNSGALTSG---VH-TFPAVL-QSSGLYSSVTVPSSSLGTOTPTTC 227
 309 ---HOENILVMRATOLQKUL-----TCEVWGPTSPKMLSLKLENKEAKVSKREKPYW 360
 228 NVNKKSNNTVDKRVELKTPLDGTTHTC---PRCP-----EPR-SCDTPPCP 271
 361 LNPBAGMOCCLSDSGOVLLESNIKVLPTWSTVPVCP-APBPSGCDKTHTC-----PELT 414
 272 RCPBP-----KSCDTPPCPCPCPCPCPCPCPCPCPCPCPCPCPCPCPCPCPC 309
 415 GGSVSLFPPKPKDTLMISRTPEVTCVVDVSHEDPEVKNMVVDGVEVNAATKPREQ 474
 310 GGSVSLFPPKPKDTLMISRTPEVTCVVDVSHEDPEVKNMVVDGVEVNAATKPREQ 369
 475 YNSTRVSVLTJLHODMLNGKEYCKVSNKALPAPIEKTIISAKGQPREPQVYTLPPSR 534
 370 FNSITFRVSVLTJLHODMLNGKEYCKVSNKALPAPIEKTIISAKGQPREPQVYTLPPSR 429
 535 DELTKQVSLTCLVKGFPSPDIAVWESNQGPPNNYKTPPVLDSDGSFFLYSKLTVDKS 594
 430 EBMTKIQVSLTCLVKGFPSPDIAVWESNQGPPNNYKTPPVLDSDGSFFLYSKLTVDKS 489

QY 595 RMOQGNVFCSSVMHEALNHYTKSLSPG 625
 Db 490 RMQGNITFCSSVMHEALNHYTKSLSPG 520

RESULT 18

Q8TC63 PRELIMINARY; PRT; 473 AA.
 ID Q8TC63;
 AC Q8TC63;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 NCBI_Taxid=9606;
 RX [1]
 RN SEQUENCE FROM N.A.
 RC TISSUE=Kidney;
 RA Strausberg R.;
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC025985; AAH25985.1; -.
 DR GO; GO:0005507; F:coppper ion binding; IEA.
 DR GO; GO:0005489; F:electron transporter activity; IEA.
 DR GO; GO:0006118; P:electron transport; IEA.
 DR InterPro; IPR000923; BlueCu_1.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003006; Ig_MHC.
 DR InterPro; IPR003596; Ig_V.
 DR Pfam; PF00047; Ig_4.
 DR SMART; SM00406; IGV; 1.
 DR PROSITE; PS00196; COPPER BLUE; 1.
 DR PROSITE; PS00835; IG_LIKE; 4.
 DR PROSITE; PS00290; IG_MHC; 3.
 DR Hypothetical protein.
 KW SEQUENCE 473 AA; 51986 MW; E29920B09BA369F5 CRC64;
 SQ
 Query Match 33.1%; Score 1130; DB 4; Length 473;
 Best Local Similarity 51.3%; Pred. No. 2.6e-78;
 Matches 260; Conservative 42; Mismatches 107; Indels 98; Gaps 15;
 Db
 138 SLTLTLESPGSSPVPVQ--CRSPKKNIOGKTLVS-----QLELDQSGTWCTVLQ 189
 45 SLTCTVSGSVASSSYVMGWQPPQKGLMWTGTFNSNMVYSPSLRBRVTMSADMSEN 104
 190 QKKVEFKIDIVLAFOKASSIYKKEGQVEFSFPLAFTVEKLTGSGELMWAERASSK 246
 105 --SFYTKLDSVTAA--DTAVVYCAAGHLV-----MGGAHMGQKLVSVSP 146
 247 -SSKSWITPDLNKKEYSVGRVQDPKLOMGKPLHLTLPLQALPQVAG--SGNTLALAE 302
 147 ASTKGPVFPPL---APCSRSTSESTAAAGCLVKQYFPPEPTVSNWSGALTSGVHTFPV 202
 303 AKTGKLEHENVLMRATOL-QKNLTCEVWGPTSPKMLSLKLENK--EAKVSKREKPYW 359
 203 LSSGLISLSSVTVYPSSSLGTIKYTCNV-----DHKPSATKDKR----- 243
 360 VINPAGMOCCLSDSGOVLLESNIKVLPTWSTVPVCP-APBPSGCDKTHTCPELIGPS 418
 244 -----VESK-----YGPCCSCPA-----PEFLGSPS 265
 419 VFLFPPKPKDTLMISRTPEVTCVVDVSHEDPEVKNMVVDGVEVNAATKPREQYNST 478
 266 VFLFPPKPKDTLMISRTPEVTCVVDVSHEDPEVKNMVVDGVEVNAATKPREQYNST 325
 479 YRVSVSVLTJLHODMLNGKEYCKVSNKALPAPIEKTIISAKGQPREPQVYTLPPSR 538
 326 YRVSVSVLTJLHODMLNGKEYCKVSNKALPAPIEKTIISAKGQPREPQVYTLPPSR 385
 539 KQVSVSLTCLVKGFPSPDIAVWESNQGPPNNYKTPPVLDSDGSFFLYSKLTVDKSRWQ 598
 386 KQVSVSLTCLVKGFPSPDIAVWESNQGPPNNYKTPPVLDSDGSFFLYSKLTVDKSRWQ 445


```

Db      356 KVVQVAVPENGMLWQCLLSBBDKXKMDSRIOVL-----SRGVNQVFLACVLG 402
Qy      416 GSGVFL 421
Db      403 GSGFPL 408

RESULT 21
ID      055054      PRELIMINARY;      PRT;      433 AA.
AC      055054;
DT      01-JUN-1998 (Tremblrel. 06, Created)
DT      01-JUN-1998 (Tremblrel. 06, Last sequence update)
DT      01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE      T4 surface glycoprotein (Fragment).
GN      CD4.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10990;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=88097446; PubMed=3501122;
RA      Madden P.J., Molineaux S.M., Madden D.E., Zimmerman K.A., Godfrey M.,
RA      Alt F.W., Chese L., Axel R.;
RT      "structure and expression of the human and mouse T4 genes.";
RL      Proc. Natl. Acad. Sci. U.S.A. 84:9155-9159(1987).
RN      [2]
RP      SEQUENCE FROM N.A.
RA      Madden P.J., Molineaux S.M., Madden D.E., Zimmerman K.A., Godfrey M.,
RA      Alt F.W., Chese L., Axel R.;
RT      Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RL      EMBL: AF045882; AAC01764.1; -.
DR      HSSP; P01730; IMR.
DR      MGD; MGI:88335; Cd4.
DR      GO; GO:0016020; C:membrane; IEA.
DR      GO; GO:006955; P:immune response; IEA.
DR      InterPro; IPR000973; CD4_TcRg.
DR      InterPro; IPR007110; Ig_Like.
DR      InterPro; IPR003596; Ig_v.
DR      Pfam; PF00047; Ig_2.
DR      PRINTS; PR00692; CD4TCANTIGEN.
DR      SMART; SM00406; IGV; 1.
DR      PROSITE; PSS0835; IG_Like; 1.
FT      NON_TER
FT      TER
SQ      SEQUENCE 433 AA; 48590 MW; AB19330750A8499A CRC64;

Query Match      28.1%; Score 958; DB 11; Length 433;
Best Local Similarity 51.0%; Pred. No. 3.8e-65;
Matches 205; Conservative 63; Mismatches 112; Indels 22; Gaps 7;

Qy      24 QGNKVVLGKKGVDELCTASQKKSIOFHKNSNOIKILGNQ--SPLTKG--PSKLNDR 80
Db      1 QGKTVLVGKGESALPCESSQKKITVFTWKFSQKRIIGQHGKGVLRGSGPSQF--DRF 59

Qy      81 DSRRLSDQGNFPLIIKNIKLIEDSDTYICEVDQKEVOLLVGLTANSTHLLQGSIL 140
Db      60 DSKKGMEKSGFPLINKLIMEDSQTYICLENRKEVELMVFVFSPTSLLOQSIL 119

Qy      141 LTHES--PGSSPVOCSPRGKNIQGGKTLVSQLELQDSGTCTCTVLOKQKVEPKID 199
Db      120 LTTDSNSKVNPLTECHGKKGKVVSGSKVLSMSNLRRQDSDFNNCTVTLQKKNMIGMTL 179

Qy      200 VVLAFAKASSIVYKKEGEQVEFSPLAFTVEKLTGSGELMWAERASSSKMTITPDLK 259
Db      180 SVLGFGSTAITAYKSEGESAFSPFLNFAE--NGMELMWMKAEKXSFQPPWISFSLK 237

Qy      260 EVSVKRVYTPDKIQMGKKLPLHLTLPOALPOVAGSNLTALDAKTKGKHQEVNLVY 319
Db      238 EVSVKSTKDLKQLKLTPLTLTKIPQVSLQFAGSGNLTLTD--KGLTHQEVNLVYMK 295

Qy      320 TOLQKULTCVWGPVTPSKMLSLKLENKAKVSKREKPVVNLPEAGMWCCLLSDSG 379

```

```

Db      296 AQLNNTICEVWGPSPKRLTLKQENQEARSESGKVVQVAVPENGMLWQCLLSBBDKXK 355
Qy      380 LESNITKVLPTMSTVPKPAPEPKSCDKHTTCCELLGSGSVFL 421
Db      356 MDSRIQVL-----SRGVNQVFLACVLGSGSPFL 384

RESULT 22
ID      095M34      PRELIMINARY;      PRT;      337 AA.
AC      095M34;
DT      01-DEC-2001 (Tremblrel. 19, Created)
DT      01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT      01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE      Immunoglobulin gamma 1 heavy chain constant region (Fragment).
GN      IGHCL.
OS      Equus caballus (Horse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX      NCBI_TaxID=9796;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Wagner B.;
RT      Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
RN      [2]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=98383416; PubMed=9717671;
RA      Wagner B., Overesch G., Sheoran A., Holmes M., Richards C.,
RA      Leibold W., Radbruch A.;
RT      "Organization of the equine immunoglobulin heavy chain constant region
RL      genes. III. Alignment of c-mu, c-gamma, c-epsilon and c-alpha genes.";
RL      Immunobiology 199;105-119(1998).
DR      EMBL; AJ300675; CAC44624.1; -.
DR      InterPro; IPR007110; Ig_Like.
DR      InterPro; IPR003597; Ig_C1.
DR      InterPro; IPR003006; Ig_MHC.
DR      Pfam; PR00047; Ig_2.
DR      SMART; SM00407; IGCL; 2.
DR      PROSITE; PSS0835; IG_Like; 3.
DR      PROSITE; PSS0290; IG_MHC; 2.
FT      NON_TER
FT      TER
SQ      SEQUENCE 337 AA; 37438 MW; A60BF2B01DEFD1F6 CRC64;

Query Match      26.0%; Score 889; DB 6; Length 337;
Best Local Similarity 60.5%; Pred. No. 5.3e-60;
Matches 170; Conservative 36; Mismatches 49; Indels 26; Gaps 5;

Qy      371 LNSDSGVLLLESNIKV-LPTMSTPV-----PCPAPEPKSCD--KTH 408
Db      56 VLOSGFYLSMVTVPASTWTSEYICNVVAAASNFKYDKRIEPIEDNHQKVCWMSKCP 115

Qy      409 TC--PELLGSPVFLPPEPKDITLMSRTPEVTCVVVDVSHEDPEVKFMYVDGVEVNA 466
Db      116 KCPAPPELLGSPVFLFPPEPKDITLMSRTPEVTCVVVDVSHEDPEVKFMYVDGVEVNA 175

Qy      467 KTKPREQVNSTYRVVSUTVLVHOMLNKEKKVKYSNLAAPLEKTSKAGQPREQ 526
Db      176 TTRPREQVNSTYRVVSUTVLVHOMLNKEKKVKYSNLAAPLEKTSKAGQPREQ 235

Qy      527 VYTLPSRRELTKNQVSLCLVKGFFPSDIAVEMESNGCP--ENNYKTPVPLDSDGSFF 584
Db      236 VYVLAHPHELKSKVSVCLVKDYRPPRLINLEMGNSNGPELETYKSTTQAQDSDGSYF 295

Qy      585 LYSKLTVDKSRVQGVNFSQVMEALNHYTQKSLSPG 625
Db      296 LYSKLTVDKSRVQGVNFSQVMEALNHYTQKSLSPG 336

RESULT 23
ID      07TMK1      PRELIMINARY;      PRT;      470 AA.
AC      07TMK1;

```

```

DT 01-OCT-2003 (Tremblrel. 25, Created)
DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CZECH II; TISSUE=Breast tumor;
RX MEDLINE=22386257; PubMed=12477932;
RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marzetta K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uedlin T.B., Toshilyuk S., Carinici P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Boeak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S.,
RA Krzywnicki M.I., Skolnik U., Smalins D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CZECH II; TISSUE=Breast tumor;
RA Strauberg R.;
RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC055910; AAB55910.1; -
KM Hypothetical protein.
SQ SEQUENCE 470 AA; 51727 MW; 6D90E4DF86B090 CRC64;

Query Match 25.0%; Score 854; DB 11; Length 470;
Best Local Similarity 38.6%; Pred. No. 4.1e-57;
Matches 203; Conservative 60; Mismatches 127; Indels 136; Gaps 15;

QY 147 PGSSPSVQCRSPRGKNIQG-----GKTLVSQLELDGSGTWGTCVLOQNKVVEFK 196
DB 33 PGASVAKISCKA-SGYTFGYMHVWKQSHGKSLKLEWIGLVNPSNGDTS---YNQK--FK 84
QY 197 IDIVLVLFQKASSIVYKKEGQVEFSPFLAFTVEKLTGSGELMWQAERASSKSNITFDL 256
DB 85 -GKATLVVDKSSSTAYNE-----LNSLTSEDSAVYYCARYYSGSYWYFDV 129
QY 257 --KNKEVSVKRVTDQPKLQMGKLPRLTLPLQALPQYASG-----NLTLALEAK----- 304
DB 130 WGAQTYTVSSAT-----TTASVYPLVPGCGDTSGSSVTTLGCLVKGTFPE 175
QY 305 -----TGKLGQEVNLV-----VMRATOLQKNLTCEWGPSTPKMLLS 341
DB 176 PVTVMKMYGALLSSGVRITVSVLQSGFYSLSLVTPSTWPSQTVCNVAAHPAS----- 229
QY 342 LKLENKAKVSKREKPPVWVLPNPAAGMOCCLSDGCVLBSNITVLTWSTPVPYCPAPRP 401
DB 230 -----RTLEIKRLEP-----RIKPPSTPP 248
QY 402 KSCDKHTCP--ELGGPSVFLPPPKDPTLMISRTPEVTCVVVDVSHEDPEVKFNMYVD 459
DB 249 GS-----SCPRGNLTGGSVFIFPPKRDALMISLTIRKTCVVVDVSEDDPDVAVHVFVD 303
QY 460 GVEVNAKTKPREQVNSTVRVSVLVTLVHODMNGKVEYCKKVSNAKLPAPIETKISKAK 519
DB 304 NKEVHTAMTQPREAQVNSTFRVVALPIQHDMMRGKEFKCKVANKKALPAPIETISKPK 363

```

```

QY 520 GQPREQVYTLPPSRDELITNOVSLTCLVKGFPYSDIAVESNGQEPENNYKTTPTPLVDS 579
DB 364 GAGTQPVYITLPPPREGMSKKKVSITCLVTNPFSEALSVEMENGELEQGYKNTPTPLDS 423
QY 580 DGSEFLYSKLTVDKSRQGGVFCGSMYMEALHNHTYQKLSLSPG 625
DB 424 DGYTFLYSKLTVDTSVLQGEIFTCSSVVEALHNHTYQKLSRSPG 469

RESULT 24
QY 09RIA4 PRELIMINARY; PRT; 437 AA.
ID 09RIA4
AC 09RIA4;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Gamma1 heavy chain of Mab7 (Fragment).
GN IGH-4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Wilde K.G., Yu X., Ekramoddoullah A.K.M., Misra S.;
RT "Cloning of cDNAs encoding for anti-white pine blister rust monoclonal
RT antibody (Mab 7, its light and heavy chains) and construction of a
RT single chain antibody (scFv).";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF152372; AAD40243.1; -
DR PIR; B45837; B45837.
DR PDB; 1COK; 11-SEP-99.
DR PDB; 1I91; 25-DEC-02.
DR PDB; 1KCU; 11-MAY-02.
DR MCD; MGI:96446; IGH-4.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; Ig_3.
DR SMART; SM00406; IgV_1.
DR PROSITE; PS00835; IG_LIKE_4.
DR PROSITE; PS00290; IG_MHC_1.
FT NON_TER 1
FT NON_TER 437
SQ SEQUENCE 437 AA; 48142 MW; 5C3A7B3E7D697C CRC64;

Query Match 24.6%; Score 841.5; DB 11; Length 437;
Best Local Similarity 33.2%; Pred. No. 3.4e-56;
Matches 200; Conservative 75; Mismatches 145; Indels 183; Gaps 14;

QY 30 LGKGDVVELTCTAS--QKKSIOFHKNSNQIKILNGQSFLLTKGSKLNDRADSRSLW 87
DB 10 LVKPGSLKLKSCAAGFTFSSYAMSWVRQTPREKLEWVASFSGGIIYYDSVYGRFTIY 69
QY 88 ----DQGNFLLINKLKEDSDTYICEVEDKEKERVQLVFLGTANSPDHLQGSGLTLE 144
DB 70 KQDKNILSLQMSLSKEDTAMTYC-----ARGDSAYWGPTLVTS 112
QY 145 SPGSSPSVQCRSPRGKNIQGKTLVSQLELDGSGTWGTCVLOQNKVVEFKIDIVLAF 204
DB 113 AAKTTTPSYPLAP-----GSAAGTNSMTL----- 138
QY 205 QKASSIVYKKEGQVEFSPFLAFTVEKLTGSGELMWQAERASSKSNITFDLKNKEVSVK 264
DB 139 -----GCLVKGFPPEPTVT-----MNSGSLSSG----- 162
QY 265 RVNQDPLQMGKLPRLTLPLQALPQ--VAGSGNLTALAEAKTKLHGQEVNLVVMRATOL 322
DB 163 -----VH--TFPAVLQSDLYTLSSVT-----VPSSTWP 189
QY 323 QKNLTCEWGPSTPKMLLSIKLENKAKVSKREKPPVWVLPNPAAGMOCCLSDGCVLLES 382
DB 190 SETVTCNVAAHPAS-----STKVDK-----IVPRDGCCKPCI----- 221

```

```
QY 383 NIKVLPWTSTPVCPCAPBPKSCDKTHTCPBLLGSGVFLPPPKKDTLMISRTPEVTCV 442
DB 222 -----CTVPEVS-----SVFIFFPKPKDVLITLTPKVTCAV 253
QY 443 VDVSHDEPEVKFMYVDGVEVHNACTKPREQYNSTYRVSVLTVLHODMLNGKEVKCY 502
DB 254 VDISKDDPEVQFSMFDDVEVHTAQTQPREQFNSTFRSVSELPIMHODMLNGKEVKCY 313
QY 503 SNKALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWES 562
DB 314 NSAAFPAPRIEKTISKTKGKPRAPQVYTIIPPEKQMAKDKVSLTCLITDFFPEDIVEMQ 373
QY 563 NGQPENNYKTPPVLDSDGSFFLYSKLTVDKSRWQOGNVPSCVMEALHNHYTQKSLS 622
DB 374 NGQPAENYKNTQPIMDTDSGYFYYSKLVNQSNWEAGNTFTCSVLHEDLNHTKLSH 433
QY 623 SPG 625
DB 434 SPG 436

RESULT 25
Q8R3V9 PRELIMINARY; PRT; 469 AA.
ID Q8R3V9 AC Q8R3V9;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN IGH-4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC024405; AAH24405.1; -.
DR PIR; B45837; B45837.
DR MGI; MGI:96446; Igh-4.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; IgV_3.
DR SMART; SM00406; IGV_1.
DR PROSITE; PS00835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 1.
KW Hypothetical protein.
SQ SEQUENCE 469 AA; 51976 MW; 534793F15D05457 CRC64;

Query Match 24.6%; Score 841.5; DB 11; Length 469;
Best Local Similarity 34.6%; Pred. No. 3.8e-56;
Matches 187; Conservative 76; Mismatches 112; Indels 165; Gaps 14;

QY 147 PGSSPSVQC-----RSFRGK-----NIQGGKTL 170
DB 33 PGSSLRSLSCAAGFTFTDYMSWRQPPGKALEMLGFRNKANGYTTESASVKGRTIS 92
QY 171 -----VQLELQDSGTWTCTVQLQNKVFEFIDIVLAFQKASSIVYKKKEGQV 219
DB 93 RDNSSQSLIYQNALRAEDSATYYCA--RDRSSSY-----YSGTSEFAWGGTLLV 141
QY 220 EFS-----FPLAFTVEKLTGSGELMWAQERASSKSWITFDLKNKEVSVKRYTOP 270
DB 142 TVSAAKTTPSVYPLA-----PGS-----AAQTNSMTLGLGVKY----- 177
QY 271 KIQMGKKLPLHLTPQALPOYAGSGNLTALAEKTKGKHQEV-----NLVWRATQLOKN 325
DB 178 -----FPEPVTVTWNSSGSLSSGVHTFPAVLQSDLYTLSSSVTPSPSTWPSQ 224
QY 326 LTCEWNGPISPKMLSLKLENKAKYSKRREKPVWVNLPEAGMQLCLSDSGQVLLSNIK 385
```

```
DB 225 VTCNVNAPHS-----STKYDK-----IVPRCCGCKPCI----- 253
QY 386 VLPWTSTPVCPCAPBPKSCDKTHTCPBLLGSGVFLPPPKKDTLMISRTPEVTCVVDV 445
DB 254 -----CTVPEVS-----SVFIFFPKPKDVLITLTPKVTCAV 288
QY 446 SHEDPEVKFMYVDGVEVHNACTKPREQYNSTYRVSVLTVLHODMLNGKEVKCYVSK 505
DB 289 SKDDEVEQFSMFDDVEVHTAQTQPREQFNSTFRSVSELPIMHODMLNGKEVKCYVNSA 348
QY 506 ALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNQ 565
DB 349 APPAPIEKTISKTKGKPRAPQVYTIIPPEKQMAKDKVSLTCLITDFFPEDIVEMQ 408
QY 566 PENNYKTPPVLDSDGSFFLYSKLTVDKSRWQOGNVPSCVMEALHNHYTQKSLS 625
DB 409 PAENYKNTQPIMDTDSGYFYYSKLVNQSNWEAGNTFTCSVLHEDLNHTKLSHSPG 468

RESULT 26
Q99LC4 PRELIMINARY; PRT; 463 AA.
ID Q99LC4 AC Q99LC4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Similar to RIKEN cDNA 181060009 gene.
GN IGH-4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC003435; AAH03435.1; -.
DR PIR; B45837; B45837.
DR HSP; P01842; 7FAB.
DR MGI; MGI:96446; Igh-4.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; IgV_3.
DR SMART; SM00406; IGV_1.
DR PROSITE; PS00835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 1.
SQ SEQUENCE 463 AA; 51007 MW; EAA674C6BBC30783 CRC64;

Query Match 24.5%; Score 835; DB 11; Length 463;
Best Local Similarity 36.5%; Pred. No. 1.2e-55;
Matches 186; Conservative 73; Mismatches 123; Indels 128; Gaps 13;

QY 136 GQSLTTLTSPSSPSVQCPSRPGK-NIQGGKTL-----VQLELQDSGTWTCTVQLN 189
DB 61 GQGLEWGEIYPSGSGNTYISEKFKGKATLTTRKSSSTAYMHLSLTSESAVYFCA----- 116
QY 190 OKKVEKIDIVLAFQKASSIVYKKKEGQVERF-----FPLAFTVEKLTGSGELMW 240
DB 117 -SSYSYSL-----FAWGGTLLVTSAAKTPPSVYPLA-----PGS----- 154
QY 241 QAERASSKSWITFDLKNKEVSVKRYTOPKIQMGKKLPLHLTPQALPOYAGSGNLT 300
DB 155 -----AAQTNSMTLGLGVKY-----FPEPVTVTWNSSGSLSSG 188
QY 301 LEAKTKGKHQEV-----NLVWRATQLOKNLTCEVWGPTSPKMLSLKLENKAKYSKE 355
DB 189 VHTFPAVLQSDLYTLSSSVTPSPSTWPSSTVTCNVNAPHS-----STKYDK- 225
QY 356 KVVWVNLPEAGMQLCLSDSGQVLLSNIKVLPWTSTPVCPCAPBPKSCDKTHTCPBLLG 415
DB 236 -----IVPRCCGCKPCI-----CTVPEVS----- 254
```

```

Qy 416 GPSVFLPFPKPKDTLMTSRTPEVTCVVVDVSHEDPEVKFNMYVDGVEVHNAKTKRPREOQ 475
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 255 --SVFIFPFPKPKDVLITLTPKTCVVVDISKDPEVGFMSFVDDVEVHTAQTQPREOQF 312
Qy 476 NSTRVVSVLTVLHODMNGEYVKCKVSKNKLPAPIEKTISKAKGQPREPVYTLPSRD 535
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 313 NSTERSVSELPIMHODMNGEYKFCRVNSAFAPIEKTISKAKGPRAPQVYTIPEPKE 372
Qy 536 ELTNQVSLTCLVNGFVPSDIAVEMESGOPENNYKTPPVLDSDGFFLYSKLTVDKSR 595
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 373 QMAQDKVSLTCLMIDTFPEPDITVEQMNGPAENYKNTQPTMDTDSYFIYSKLVNOKSN 432
Qy 596 WQGNVFSQSVHMEALHNHYTKSLSPG 625
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 433 WEAGNTFCVSLHNEGILHNHTEKSLSPG 462

RESULT 27
099L31 PRELIMINARY; PRT; 468 AA.
AC 099L31;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Similar to RIKEN cDNA 1810060009 gene.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Straube R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC003878; AA03878.1; -.
DR PDB; 2AP2; 24-NOV-99.
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; IG_3.
DR SMART; SM00406; IGv_1.
DR PROSITE; PSS0835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 1.
SQ SEQUENCE 468 AA; 51661 MW; 96352328B332ADB CRC64;

Query March 24.0%; Score 819; DB 11; Length 468;
Best Local Similarity 35.1%; Pred. No. 2e-54;
Matches 214; Conservative 62; Mismatches 148; Indels 186; Gaps 19;
Qy 30 LGKGGDTVELTCTAS--QKSIQFMKNSNGIKLNGQ-----SFLTKGPRSKLN 77
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 30 LVKRGASVKLSCTASGKFNKQSLMHW-----VKORPEGLWICWIDPEDGETYARFQ 84
Qy 78 DRADRSRLMDQGNFPIIKNLKIEDSDTYICEFDQKEEYQLVFGILTANSDFHLQOQ 137
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 85 DKATITDTSNTAV-LQLSLTSEDTAIYC-----ARNLLYG-----GYDVMQO 130
Qy 138 SLTITLSPGSSPSVQCRSRGNKIQGKTLVSQELQDSGTWCTCTVLQNKQKVEFKI 197
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 131 GTTITVSAAKTATASVPLAV-----VCGDIT----- 157
Qy 198 DIVVLQAKASSIYKKEGEVERSPFLATVEKLTGSGELMWAERASSKSWITDLEK 257
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 158 -----GSSVTL-----GCLVKYFP-----EPTV-----LTWNSGLSSG----- 187
Qy 258 NKEVSVKRVTPDPLQNGKKLPLHLTLPOALPO--VAGSGNLTALBAKTGKLGHEVNLV 315
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 188 -----VA--TFPAVLQSDLYLTSSSVT----- 207
Qy 316 VMRAITOLQKNTCEVWGFTSPKMLSLKLENKAKVSRKRPVAVLNPEAGMQCLLSDS 375
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 208 VTSSTWSPQSITCVAVHAPS-----STKVDK----- 233

```

```

Qy 376 GQVLLBSNIVKLPWSTPVPCEPAPEPKSCDKHTHCELLGSPVFLPFPKDTLMTSRT 435
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 234 -----KIEPRGPIIKCP---PKC-----PAPVLLGGPSVFIRPKIKQVLMISLS 277
Qy 436 PEVTCVVVDVSHEDPEVKFNMYVDGVEVHNAKTKRPREOYNSTYRVVSVLTVLHODMNG 495
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 278 PMVTCVVVDVSEDDPDVQISMFVNNVSVLTAQTQTHREDYNSITLVSAALPIQHDMMWG 337
Qy 496 KEYKCKSNKALPAPIEKTISKAKGQPREPVYTLPSRDRELTNQVSLTCLVNGFVPSD 555
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 338 KEFKCKVNNKALPAPIERTISKPKGSYRAQVYVLLPPEEEMTKKQVTLTCMTVDPEMD 397
Qy 556 IAVEMESNGOPENNYKTPPVLDSDGFFLYSKLTVDKSRWQGNVFSQSVHMEALHNHY 615
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 398 IYVEMTNNGKTELNTKTEPEVLDSDGFFLYSKLTVDKSRWQGNVFSQSVHMEALHNH 457
Qy 616 TOKSLSLSPG 625
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 458 TTKSFSRTPG 467

RESULT 28
09DBL4 PRELIMINARY; PRT; 473 AA.
AC 09DBL4;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE 1810060009Rik protein.
DE IGH-1 OR 1810060009RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=C57BL/6J; TISSUE=Pancreas;
RX MEDLINE=21085660; Pubmed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleisemann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirml L.M., Staudt P., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Guelastic S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombarts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilmink L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohzuki S.,
RA Hayashizaki Y.,
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK007918; BAB25349.1; -.
DR PIR; S26746; S26746.
DR HSSP; P01842; TPAB.
DR MGD; MGI:96443; Igh-1.
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; IG_3.
DR SMART; SM00406; IGv_1.
DR PROSITE; PSS0835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 1.
SQ SEQUENCE 473 AA; 51699 MW; 9DE057A514475FBB CRC64;

Query March 23.6%; Score 805.5; DB 11; Length 473;
Best Local Similarity 34.0%; Pred. No. 2.2e-53;

```



```

Db 90 S-RDNAKTFLF-----LQMTSLSESDTAMYYCARELMLRRID-----YWG 128
Qy 137 QSLTLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELQDSGTWTCTVLQNGKVEFK 196
Db 129 OCTITIVSSAKTTPSVYPLAPGCCDTTG----- 157
Qy 197 IDIVVLAFQKASIVYKKEGEOVESFPLAFTVEKLTGSGELMWAQERASSKSWITFDL 256
Db 158 -----SSVTL---GCLVKGYPFESVTVT-----MNSGLSSS----- 186
Qy 257 KKEKESVVRVYQDPLQMGKLLPLHLTLPOLPOYAGSGNLTALAEKTKLHDEVLV 316
Db 187 -----VH-TFPALL-----QSGLYTWMSSSVTV 207
Qy 317 MRATOLQKLTCEVWGPTSPKMLSLKENKAQVSKREKPVWVLPNPAQMGCLLSNSG 376
Db 208 PSSITWPSQTVTCSVAHPAS-----STTVDKKLEP-----SG 238
Qy 377 QVLESNIKVLPTWSTPVPCEPAPEPKSCDKTHTC--PELLGSPSVFLFPKPKDTLMISR 434
Db 239 PI-----STINPCF-----PCKECHKCPAPLBEGSPVFIIPPIKQVLMISL 281
Qy 435 TPETVTCVVVDVSHEDPEVKFMYDGVENNAKTRPREQYNSTYRVVSVLTVLIHOMLN 494
Db 282 TPKVTCTVVVDVSEDDPDVQISMFWNNVEVHTAQTQTHREDYNSTIRVVSALPIQHODWMS 341
Qy 495 GKEYKCKVSNKALPAPIEKTIISKAKGPREPOVYTLPPSRDELTKNQVSLTCLVKGFYPS 554
Db 342 GKEKCKVNNKQDLSPIERTISKIKGLVRAPOVYILPPRAQSLRKQDSLCLVAGFNPQ 401
Qy 555 DIAEWESNGQPNENNYKTPPVLDSDGSFFLYSKLTVDKSRMQGNVSCVMHEALHNH 614
Db 402 DISYEMWSNGHTEENYKDTAPVLDSDGSFYISKLDITKTSKMEKTDSDSCVNRHEGLKNY 461
Qy 615 YTKQSLSLSPG 625
Db 462 YLKKTISRSPG 472

```

RESULT 31

```

Qy 09R3H6 PRELIMINARY; PRT; 474 AA.
AC 09R3H6;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Hypothetical protein.
GN A004919.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strauberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC025447; AAH25447.1; -
DR MGI; MGI:2144967; A004919.
DR GO; GO:0005489; F:electron transporter activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR000345; Cytochrome_c;
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; Ig; 3.
DR SMART; SM00406; IgV; 1.
DR PROSITE; PS00190; CYTOCHROME_C; 1.
DR PROSITE; PS00835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 1.
DR PROSITE; PS00047; Ig_v.
DR SMART; SM00406; IgV; 1.
DR PROSITE; PS00835; IG_LIKE; 5.
Qy SEQUENCE 474 AA; 51748 MW; 8608B576CD2874A CRC64;

```

Query Match 23.3%; Score 794.5; DB 11; Length 474;

Best Local Similarity 33.4%; Pred. No. 1.6e-52;
Matches 211; Conservative 72; Mismatches 150; Indels 199; Gaps 20;

```

Qy 15 QIALPPAATQGNKVLNKKGDTVELTCTASQ--KKSIOFWKQ-----SNQIKILCN 64
Db 20 QVOLLQSGPE---LVKPGASVKISCRAGYAPRSK---MNMVKKRRPQKGLIEWIGRI 70
Qy 65 -----QGSFLTKGPKSLNDRADRSLSMDGPNPLIINKIKIDSDTIDYIEVEDQK 115
Db 71 FPGDGDTHYSGKF--QKAKLTADKSSVTF-----LQTLTSEDSAVYPCARSD 120
Qy 116 EEVQLVFGLTANSDTHLQGSQSLTLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLE 175
Db 121 -----YG-----DYDDMGQGATVTVSSAKTTPSVYPLAPGCCDTTG----- 158
Qy 176 LQDSGTWCTVLQNGKVEFKIDIVLAFQKASIVYKKEGEOVESFPLAFTVEKLTGSG 235
Db 159 -----SSVTL---GCLVKGYPFESVTVT----- 178
Qy 236 GELMWAERASSSKSWITFDLKNKEVSVKRVTQDPLQMGKLLPLHLTLPOLPOYAGSG 295
Db 179 -----MNSGLSSS-----VH-TFPALL----- 195
Qy 296 NLTLAEAKTGLKHOEVLVYMRATOLQKLTCEVWGPTSPKMLSLKENKAQVSKRE 355
Db 196 -----QSGLYTWMSSSVTVPSSTWPSQTVTCSVAHPAS-----STTVDKL 235
Qy 356 KPVWVLPNPAQMGCLLSDSGQVLESNIKVLPTWSTPVPCEPAPEPKSCDKTHTC--PEL 413
Db 236 EP-----SGPI-----STINPCF-----PCKECHKCPAPNL 261
Qy 414 LGGSVFLFPKPKDTLMISRTPETVTCVVVDVSHEDPEVKFMYDGVENNAKTRPRE 473
Db 262 EGGSPVFIIPPIKQVLMISLTPKVTCTVVVDVSEDDPDVQISMFWNNVEVHTAQTQTHRE 321
Qy 474 QYNSTYRVVSVLTVLIHOMLNKQVSLTCLVKGFYPSKLTVDKSRMQGNVSCVMHEALHNH 533
Db 322 DYNSTIRVVSALPIQHODWMSGKEFKKNNKQDLSPIERTISKIKGLVRAPOVYILPPP 381
Qy 534 RDELTKNQVSLTCLVKGFYPSDIAEWESNGQPNENNYKTPPVLDSDGSFFLYSKLTVDK 593
Db 382 AEOLSRKQDLSLTCLVAGFNPQDISVETWSNGHTEENYKDTAPVLDSDGSFYISKLDIKT 441
Qy 594 SRMQGNVSCVMHEALHNHNYTKQSLSLSPG 625
Db 442 SKMEKTDSDSCVNRHEGLKNYILKKTISRSPG 473

```

RESULT 32

```

Qy 096EYO PRELIMINARY; PRT; 613 AA.
AC 096EYO;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Strauberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC011857; AAH11857.1; -
DR PIR; S15590; S15590.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; Ig; 5.
DR SMART; SM00406; IgV; 1.
DR PROSITE; PS00835; IG_LIKE; 5.

```

DR PROSITE; PS00290; IG_MHC; 3.
KW Hypothetical protein.
SQ SEQUENCE 613 AA; 67273 MW; 31214203FB8421E7 CRC64;

```

DT 01-OCT-2003 (Tremblrel. 25, last annotation update)
DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/LEU-3)
DE (Fragment).
OS Mandrillae sphinx (Mandrill) (Papio sphinx).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopitheciinae; Mandrillae.
OX NCBI_TaxID=9561;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98320644; PubMed=9656488;
RA Harris E.B., Disotell T.R.;
RT "Nuclear gene trees and the phylogenetic relationships of the
RT mangabeys (primates: Papionini).";
RL Mol. Biol. Evol. 15:892-900(1998).
CC -1- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL
CC RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY
CC SIMILARITY).
CC -1- SUBUNIT: ASSOCIATES WITH P56-LCK (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC ONE V-LIKE AND 1 C2-LIKE DOMAIN.
DR EMBL; AF057386; AAC25130.1; -.
DR HSSP; P01730; 1CDY.
DR GO; GO:0016021; C:Integral to membrane; IEA.
DR InterPro; IPR007110; IG-like.
KM Immunoglobulin domain; T-cell; MHC; Transmembrane.
FT NON_TER 1
FT DOMAIN <1 19 IG-LIKE V-TYPE DOMAIN.
FT DISULFD 20 >86 IG-LIKE C2-TYPE DOMAIN.
FT NON_TER 49 78 BY SIMILARITY.
FT SEQUENCE 86 AA; 9406 MW; 2BD97A9BE19582AB CRC64;
SQ
Query Match 11.7%; Score 400; DB 6; Length 86;
Best Local Similarity 89.4%; Pred. No. 2.7e-23;
Matches 76; Conservative 6; Mismatches 3; Indels 0; Gaps 0;
OY 107 YICVEDEKKEVGLVGLTANSDTHLQGSSTLTLESPPGSSPSVQCRPRGKNIQGG 166
DB 1 YICVEDEKKEVGLVGLTANSDTHLQGSSTLTLESPPGSSPSVQCRPRGKNIQGG 60
OY 167 KTLVSQLELDQSGTWTCTVLOK 191
DB 61 RTLSVPLQRQDSGTWCTVTSQDOK 85
RESULT 35
ID 077597 PRELIMINARY; PRT; 86 AA.
AC 077597;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, last annotation update)
DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/LEU-3)
DE (Fragment).
OS Mandrillae leucophaeus (Drill) (Papio leucophaeus).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopitheciinae; Mandrillae.
OX NCBI_TaxID=9568;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98320644; PubMed=9656488;
RA Harris E.B., Disotell T.R.;
RT "Nuclear gene trees and the phylogenetic relationships of the
RT mangabeys (primates: Papionini).";
RL Mol. Biol. Evol. 15:892-900(1998).
CC -1- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL
CC RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY
CC SIMILARITY).
CC -1- SUBUNIT: ASSOCIATES WITH P56-LCK (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).

```

```

CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC ONE V-LIKE AND 1 C2-LIKE DOMAIN.
DR EMBL; AF057387; AAC25131.1; -.
DR HSSP; P01730; 1CDY.
DR GO; GO:0016021; C:Integral to membrane; IEA.
DR InterPro; IPR007110; IG-like.
KM Immunoglobulin domain; T-cell; MHC; Transmembrane.
FT NON_TER 1
FT DOMAIN <1 19 IG-LIKE V-TYPE DOMAIN.
FT DISULFD 20 >86 IG-LIKE C2-TYPE DOMAIN.
FT NON_TER 49 78 BY SIMILARITY.
FT SEQUENCE 86 AA; 9406 MW; 2BD97A9BE19582AB CRC64;
SQ
Query Match 11.7%; Score 400; DB 6; Length 86;
Best Local Similarity 89.4%; Pred. No. 2.7e-23;
Matches 76; Conservative 6; Mismatches 3; Indels 0; Gaps 0;
OY 107 YICVEDEKKEVGLVGLTANSDTHLQGSSTLTLESPPGSSPSVQCRPRGKNIQGG 166
DB 1 YICVEDEKKEVGLVGLTANSDTHLQGSSTLTLESPPGSSPSVQCRPRGKNIQGG 60
OY 167 KTLVSQLELDQSGTWTCTVLOK 191
DB 61 RTLSVPLQRQDSGTWCTVTSQDOK 85
RESULT 36
ID 077MT6 PRELIMINARY; PRT; 614 AA.
AC 077MT6;
DT 01-OCT-2003 (Tremblrel. 25, Created)
DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, last annotation update)
DE Hypothetical protein.
DE Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6NCR; TISSUE=Hematopoietic Stem Cell;
RX MEDLINE=22388257; PubMed=12477937;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uebin T.B., Tomihata S., Carninci P., Prange C.,
RA Rana S.S., Loggiano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smillie D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6NCR; TISSUE=Hematopoietic Stem Cell;
RA Strausberg R.,
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC053409; AAH53409.1; -.
KM Hypothetical protein.
SQ SEQUENCE 614 AA; 67746 MW; 839BAF3B8D124F89 CRC64;
Query Match 11.7%; Score 400; DB 11; Length 614;

```

Best Local Similarity 22.5%; Pred. No. 4,6e-22;
Matches 156; Conservative 98; Mismatches 232; Indels 206; Gaps 28;

```

QY 30 LGKKGDTVELTCTAOKKSIQFMKN-----SNQIKILGN-----QGSFLTKGP 73
DB 30 LVPKASVATKSCAS--GYAFSSSMNMWVKORPGKIGRIVYPGDGTNYNGKF--KKG 86
QY 74 SKLNN-DRAASRSLMOMQGNFPLIKLKLEDSITYICEVEDKKEEVLVFGILTANSDH 132
DB 87 ATLTDKSSSTAYMQ-----LSSLTSEDSAVYFC-ARDGSSYRFAYWG----- 129
QY 133 LLOGGSLTLTLLESP--PGSSPSVQCRSP-RGNKIQGKTLVSQLELODGSFTWCTVLON 189
DB 130 --OGTIVTASBSQSPFNVPFLVSCSPSLSDKULVAMGLARPLSTISFTWN----- 181
QY 190 QKKVEPKIDIVLAFQKASSIYVKEGEQVFSFPLAFVTEKLTGSGELMWMQARASSK 249
DB 182 -----YQNNAEVYIQ-----GIR 193
QY 250 SWITPDLKNKESVKNVTDPRKQM-----GKLLPLHLPLPOLPYAGSG 295
DB 194 TPELTATGKYLATSGVLLSPKSLLEGSEYLVCKIHGKKNKDLHVPILP----- 243
QY 296 NLTLALEAKTKLHOENLVV-----MRATQLOKNLTCEVMGPTSPKMLSLKLENKA 349
DB 244 -----AVAEKNPVNVVPPRDPGSGAPKPKSLICEATNFTKPTIVSWLXKQKLV 295
QY 350 KVSREKPVWVNLN---PEA-----GMQCLISDSGOVLESNIKVL 388
DB 296 ESGFTTDPVTIEBKSGTPQYVKVISTLTISEIDMLNLYTCRVDHRLGFLK----- 348
QY 389 TWSTPVPKAPREKSCDKHTCEPLGGS---VLFPRKPKDTLMISTPRTVCVVD 444
DB 349 -----NVSSSTC---AASPTDILFTIPSPAD-TFLSKSANLTLCLVSN 388
QY 445 VSHEDDEVKNMYVDGVEVNAKTKPREEQNSTYRVSVLYTLHODMLNGEYKCKVSN 504
DB 389 LATYE-TLNIWMSQSGEPLETIKIMESHPNCTFFAKGVASVCVEDMNRKKEFVCTVH 447
QY 505 KALPADIETKISAKQPRE---POVYTLPPSRDEL-T-KNOVSLTCLVKGFPYSDIAV 558
DB 448 RDLPSQKPKISK---PNEVHKGHPAVVYLLPPAREQNLRESATVCLVKGKSPADISV 503
QY 559 EWSNQ--PENNYKTPPLD--SDGSFELYKLTVDKSRMOQGVNFGSVWHEHLHN 614
DB 504 QWLQRQQLPQEKYVTSAPMPEGAPGFYTHSLTVTEEMNSGETTYTCVSHALPHL 563
QY 615 YTKSLSLSPGLQDTECAADGDELGLMTT 646
DB 564 VTERTVDS-----TGEVNAEPEGPENLMTT 590

RESULT 37
077594 PRELIMINARY; PRT; 86 AA.
ID 077594;
AC 077594;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/LEU-3) (Fragment).
OS Cercopithecus mitis.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Cercopithecus.
OC NCBI_TaxID=36225;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98320644; PubMed=9656488;
RA Harris E.E., Disotell T.R.;
RT "Nuclear gene trees and the phylogenetic relationships of the
RT mangabeys (primates: Papionini).";
RL Mol. Biol. Evol. 15:892-900(1998).

```

```

CC -1- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL
CC RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY
CC SIMILARITY).
CC -1- SUBUNIT: ASSOCIATES WITH P56-LCK (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC ONE V-LIKE AND 1 C2-LIKE DOMAIN.
DR EMBL: AF057381; AAC25125.1; -.
DR HSSP: P01730; 1CDY.
DR GO: GO:0016021; C:integral to membrane; IEA.
DR InterPro: IPR007110; I9-like.
KW Immunoglobulin domain; T-cell; MHC; Transmembrane.
FT DOMAIN 1 1
FT 1 19 IG-LIKE V-TYPE DOMAIN.
FT 20 86 IG-LIKE C2-TYPE DOMAIN.
FT DISULFID 49 78 BY SIMILARITY.
FT NON TER 86
SQ SEQUENCE 86 AA; 9417 MW; 371CA39EF58182AB CRC64;

Query Match 11.6%; Score 397; DB 6; Length 86;
Best Local Similarity 89.4%; Pred. No. 4,6e-23;
Matches 76; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 107 YICEVEDKKEEVLVFGILTANSDTHLLOGSLTLTLLESPGSSPSVQCRSPRGKNIQGG 166
DB 1 YICEVEDKKEEVLVFGILTANSDTHLLOGSLTLTLLESPGSSPSVQCRSPRGKNIQGG 60
QY 167 KTLVSQLELDQSGTWTCTVLDNOK 191
DB 61 RLISVQLERODSGTWTCTVSQDOK 85

RESULT 38
077599 PRELIMINARY; PRT; 86 AA.
ID 077599;
AC 077599;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/LEU-3) (Fragment).
OS Theropithecus gelada (Gelada baboon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Theropithecus.
OC NCBI_TaxID=9565;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98320644; PubMed=9656488;
RA Harris E.E., Disotell T.R.;
RT "Nuclear gene trees and the phylogenetic relationships of the
RT mangabeys (primates: Papionini).";
RL Mol. Biol. Evol. 15:892-900(1998).
DE -1- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL
DE RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY
DE SIMILARITY).
CC -1- SUBUNIT: ASSOCIATES WITH P56-LCK (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC ONE V-LIKE AND 1 C2-LIKE DOMAIN.
DR EMBL: AF057389; AAC25133.1; -.
DR HSSP: P01730; 1CDY.
DR GO: GO:0016021; C:integral to membrane; IEA.
DR InterPro: IPR007110; I9-like.
KW Immunoglobulin domain; T-cell; MHC; Transmembrane.
FT DOMAIN 1 1
FT 1 19 IG-LIKE V-TYPE DOMAIN.
FT 20 86 IG-LIKE C2-TYPE DOMAIN.
FT DISULFID 49 78 BY SIMILARITY.
FT NON TER 86
SQ SEQUENCE 86 AA; 9420 MW; 2BD97A98574EE9AB CRC64;

Query Match 11.6%; Score 397; DB 6; Length 86;

```

Best Local Similarity 88.2%; Pred. No. 4.6e-23;
Matches 75; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

QY 107 YICEVEDQKEEVOLLVFGLTANSPTHLQGSLLTLESPPGSSPSVQCSPPRKNIQGG 166
DB 1 YICEVEDKKEVELLVFGLTANSPTHLQGSLLTLESPPGSSPSVQCSPPRKNIQGG 60

QY 167 KTLSVSOLELODSTGWTCTVLQONK 191
DB 61 RTLSVPQLERODSGTWTCTVNSQDOK 85

RESULT 39

ID 077595 PRELIMINARY; PRT; 86 AA.
AC 077595;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/LEU-3) (Fragment).
OS Cercopithecus galericulus chrysogaester.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Cercopithecidae.
OX NCBI_TaxID=75569;

RP SEQUENCE FROM N.A.
RX MEDLINE=98320644; PubMed=9656488;
RA Harris E.E., Disotell T.R.;
RT "Nuclear gene trees and the phylogenetic relationships of the mangabeys (primates: Papionini).";
RL Mol. Biol. Evol. 15:892-900(1998).
CC -1- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY SIMILARITY).

CC -1- SUBUNIT: ASSOCIATES WITH P56-LCK (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS ONE V-LIKE AND 1 C2-LIKE DOMAIN.

DR EMBL; AF057382; AAC25126.1; -.
DR HSSP; P01730; 1CDY.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR InterPro; IPR007110; Ig-like.
KW Immunoglobulin domain; Glycoprotein; T-cell; MHC; Transmembrane.

FT DOMAIN 1 19 IG-LIKE V-TYPE DOMAIN.
FT DOMAIN 20 >86 IG-LIKE C2-TYPE DOMAIN.
FT CARBOHYD 79 79 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DISULFID 49 78 BY SIMILARITY.
FT NON_TER 86 86
SQ SEQUENCE 86 AA; 9419 MW; A9D97A98E19582BE CRC64;

Query Match 11.6%; Score 395; DB 6; Length 86;
Best Local Similarity 88.2%; Pred. No. 6.6e-23;
Matches 75; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 107 YICEVEDQKEEVOLLVFGLTANSPTHLQGSLLTLESPPGSSPSVQCSPPRKNIQGG 166
DB 1 YICEVEDKKEVELLVFGLTANSPTHLQGSLLTLESPPGSSPSVQCSPPRKNIQGG 60

QY 167 KTLSVSOLELODSTGWTCTVLQONK 191
DB 61 RTLSVPQLERODSGTWTCTVNSQDOK 85

RESULT 40

ID 077598 PRELIMINARY; PRT; 86 AA.
AC 077598;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)

DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/LEU-3) (Fragment).

OS Papio sp. (baboon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Papio.
OX NCBI_TaxID=61183;

RP SEQUENCE FROM N.A.
RX MEDLINE=98320644; PubMed=9656488;
RA Harris E.E., Disotell T.R.;

RT "Nuclear gene trees and the phylogenetic relationships of the mangabeys (primates: Papionini).";
RL Mol. Biol. Evol. 15:892-900(1998).
CC -1- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY SIMILARITY).

CC -1- SUBUNIT: ASSOCIATES WITH P56-LCK (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS ONE V-LIKE AND 1 C2-LIKE DOMAIN.

DR EMBL; AF057388; AAC25132.1; -.
DR HSSP; P01730; 1CDY.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR InterPro; IPR007110; Ig-like.
KW Immunoglobulin domain; Glycoprotein; T-cell; MHC; Transmembrane.

FT DOMAIN 1 19 IG-LIKE V-TYPE DOMAIN.
FT DOMAIN 20 >86 IG-LIKE C2-TYPE DOMAIN.
FT CARBOHYD 79 79 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DISULFID 49 78 BY SIMILARITY.
FT NON_TER 86 86
SQ SEQUENCE 86 AA; 9433 MW; A9D97A98574EE9BE CRC64;

Query Match 11.5%; Score 392; DB 6; Length 86;
Best Local Similarity 87.1%; Pred. No. 1.1e-22;
Matches 74; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

QY 107 YICEVEDQKEEVOLLVFGLTANSPTHLQGSLLTLESPPGSSPSVQCSPPRKNIQGG 166
DB 1 YICEVEDKKEVELLVFGLTANSPTHLQGSLLTLESPPGSSPSVQCSPPRKNIQGG 60

QY 167 KTLSVSOLELODSTGWTCTVLQONK 191
DB 61 RTLSVPQLERODSGTWTCTVNSQDOK 85

RESULT 41

ID 096BB9 PRELIMINARY; PRT; 597 AA.
AC 096BB9;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=B-cell;
RA Strausberg R.;
RL EMBL; BC015760; AAH15760.1; -.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig; 5.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PSS0835; IG_LIKE; 5.
DR PROSITE; PS00290; IG_MHC; 3.
KW Hypothetical protein.

SEQUENCE 597 AA; 65039 MW; 4FCA3AD8ECE263D9 CRC64;
 Query Match 11.5%; Score 391.5; DB 4; Length 597;
 Best Local Similarity 22.4%; Pred. No. 2e-21;
 Matches 143; Conservative 116; Mismatches 248; Indels 131; Gaps 24;

30 LCKKGDVTLTCTAS--QKSKIQFMKNSNQIK-----ILGNQSGFL---TKGPSLT 76
 30 LVQPGSGSLKSCASFSFSSSYAMNVRQAPGKLEWVAISGSGSYAYASVKKRFTI 89
 77 NDRADRSRLMDQGNFPLIINKIKIDSDTYICEVEDQKEVQVLTAFGLTANSDTHTLQ 136
 90 S-RDNRSDTLVQLQMN-----SLRADTAVYCAKDPRG-----YASAGVTREDYWG 135
 137 QSLTTLTLESPGSSPS---VQCRSPRGKNIQCGKTLVSQLELDQSGTWTCTVLONQK 192
 136 QGLTVLVSSGSAAPLTFPLVSC-----ENSPDSTSSVAAGCL 174
 193 VERKIDIVLAFQ-KAISTYKKEGQVERSPPLATVEKLTGSGELMMQAEASSSKSM 251
 175 QDLPLSITFSWKYKNSDISSTRG-----FVSULRGCKYAAATSOVLLPSKDV----- 222
 252 IFEDLNKEVSVKRVTDPRLOMGKYLPLHL--TLQALPOYA-----GSGNLTALAE 302
 223 ---MGSTDEHVCKVQHPRGNKEKQVPLPVIAELPRKVSFVPRDGFPGNPKSKLJC 278
 303 AKTGKLNQEVNLVVMR-----ATQLQKLTCEVWGPTSPRYMLSLKLENKEAVSKR 354
 279 QATGFSPRQIOVSWLREKQVSGVTTDQVQAEAKESGPTTYKVTSTLTIKESD----- 332
 355 EKVWVLANEAGWMOCLLSDSQVLE-SNIKVLPWMSPTVPVPRAPRPSKCDTHTCP 413
 333 ---WL---SQSMFTRCDVHRGLTFQONASSMCVPODTHI----- 366
 414 LGSPVFLPFPKPKDMLTSTRPEVTCVVVDVSHEDPEVKFNNYVDGVEVNAKTRPRE 473
 367 ---RVFAIPPS-FASIFLTKSKTKLCLVLTLDLTYD-SVTISWTRNGEAVKHTMIS 420
 474 QVNSTYRVVSVLTVLHODMLNGEKYCKKSNKALPAPIETIKTSKAGQ-REQVYTLTP 532
 421 HPNATSAVGEASICEEDMNSGERFCTVHTDLPSPKQITISRPQGVNLHREDVLLRP 480
 533 SRDEL-KNQVSLTCLVKGRYPSDIAVEMESNGQ--ENNYKTTTPVLD--SDPSFLTS 587
 481 ARQQLMRRESATITCLVTGSPADVFVQMQRQPLSPREKVTYSAPMPEQALGRIFAS 540
 588 KLTVDKSRMQGVNFSQVMEALHNHYTQKSLSPG 625
 541 ILTVSEEWMTGETYTCVVAHVALPKRVTERVTVDKSTG 578

RESULT 42
 ID 077601 PRELIMINARY; PRT; 86 AA.
 AC 077601; 01-NOV-1998 (TREMblrel. 08, Created)
 DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
 DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)
 DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/LEU-3) (Fragment).
 OC Lophoceros albigena albigena.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
 OC Cercopitheciae; Lophoceros.
 OX NCBI_Taxid=75568;
 RN SEQUENCE FROM N.A.
 RP MEDLINE=98320644; Pubmed=9656488;
 RA Harris E.E., Disotell T.R.;
 RT "Nuclear gene trees and the phylogenetic relationships of the
 RL Mol. Biol. Evol. 15:892-900(1998).
 CC -!- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL

RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY SIMILARITY).
 CC -!- SUBUNIT: ASSOCIATES WITH P56-LCK (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE IMMNOGLOBULIN SUPERFAMILY. CONTAINS ONE V-LIKE AND 1 C2-LIKE DOMAIN.
 DR HSBP; AF057391; AAC25135.1; -.
 DR HSBP; P01730; ICDY.
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR InterPro; IPR007110; I9-like.
 KW Immunoglobulin domain; T-cell; MHC; Transmembrane.
 FT NON_TER 1 1
 FT DOMAIN <1 19 IG-LIKE V-TYPE DOMAIN.
 FT DOMAIN 20 >86 IG-LIKE C2-TYPE DOMAIN.
 FT DISULFID 49 78 BY SIMILARITY.
 FT NON_TER 86 86
 SQ SEQUENCE 86 AA; 9463 MW; 2BD97A8464FE9AB CRC64;

Query Match 11.4%; Score 388; DB 6; Length 86;
 Best Local Similarity 87.1%; Pred. No. 2.3e-22;
 Matches 74; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

107 YICEVEDQKEVQVLTAFGLTANSDTHTLQGSGLTTLSPGSSPSVQCRSPRGKNIQCG 166
 1 YICEVEDKKEVEVLTAFGLTANSDTHTLQGSGLTTLSPGSPETSPSVKCRSPRGKNIQVG 60
 167 KTLVSQLELDQSGTWTCTVLONQK 191
 61 RTLVSQLELDQSGTWTCTVSQDK 85

RESULT 43
 ID 096G46 PRELIMINARY; PRT; 614 AA.
 AC 096G46; 01-DEC-2001 (TREMblrel. 19, Created)
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
 DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)
 DE Hypothetical protein.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Tissue-B-cell;
 RC Strausberg R.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC009851; AA009851.1; -.
 DR PIR; S15590; S15590.
 DR GO; GO:0005622; C:intracellular; IEA.
 DR GO; GO:0003700; F:transcription factor activity; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR000005; HTHAFC.
 DR InterPro; IPR007110; I9-like.
 DR InterPro; IPR003006; IG-MHC.
 DR InterPro; IPR003596; IGTV.
 DR InterPro; IPR00477; I9V.
 DR SMART; SM00406; IGV; 1.
 DR PROSITE; PS00041; HTH_ARAC_FAMILY_1; 1.
 DR PROSITE; PS50835; IG_LIKE; 5.
 DR PROSITE; PS00290; IG_MHC; 3.
 KW Hypothetical protein.
 SQ SEQUENCE 614 AA; 67921 MW; 55EF536E77AA9B8B CRC64;

Query Match 11.4%; Score 388; DB 4; Length 614;
 Best Local Similarity 22.0%; Pred. No. 3.8e-21;
 Matches 150; Conservative 116; Mismatches 272; Indels 144; Gaps 26;

11 LVLQALALPAATQGNKVLG---KKGDTVELTCTASQKSI--QFMKNSQIKL-- 62
 7 ILFLVAADAVASQMQVQSGAEVKKTKGSSVSKVSCRAAGVTFTYRYLHWVROAPGAL 66

```

QY 63 -----GNQGSFLTKGPSKLNDRADSRSLMDQGNFLLIKNLKIEDSDTYICEVEDQ 114
DB 67 MGWITPFGN-----TNYAQKFDQRVITTRDR-SMNTAYMELSLRSRSDTAMYYC----- 115
QY 115 KEVQLVFLGLTAN-SDTHLQGSLLTLESPPGSSPS-----VQCRSPRKNIOGKCTL 169
DB 116 -----ARGSSSWMDAFDIMGQGTWTVSSGSASAPTLPLVSC----- 154
QY 170 SVSQLELDGSGTWTCTVLQNGKVEFKIDIVLAFQ-KASSIVKKEGEVFPPLAFT 228
DB 155 -----ENSPSDTSSVAVGLAQDFLPDSITFSKWKYNNSDISTSTR-----PPSVLR 201
QY 229 VEKLTGSGELMWQAEKASSSSKSTWTFDLKNEVSKRYTQDPKLOMGKPLHL--TLFQ 286
DB 202 GGTAAATGQVLLPBGDV-----MGTDENHYCKVQHPNKNKENVPLPIALP 251
QY 287 ALPOYA-----GSGNLTALAKTKLHQLHGVNLYVNR-----ATOLQKNLTCEVW 331
DB 252 KVSFVFPDRDGFNPSPKSKLICATGFSPRQIQVSWLREGKQVSGVTTDQVQAEAKES 311
QY 332 GPTSPKMLSLKLNKKEAKVSKREKVPVVLNPEAGMQLSDSGQVILE-SNIKVLPW 390
DB 312 GPTTYKVTSTLTIKESD-----WL---SQSMFTCRVDHRLGLTFQNASMCVDPQ 358
QY 391 STVPCEPAPBPKSCDKHTTCPELLGSPVFLFPKPKDITLMSRTEPVTCVVDVSHEDP 450
DB 359 DTAI-----RVFAIPPS-FASITLTYSTKLTCLVTDLITYD- 393
QY 451 EVKENWYVDGVEVHNATKPREEOYNSTYRVSVLTVLHODLNGKEKCVSKALPAP 510
DB 394 SVTISMRQNGEAVKVTNINISHPNATFSAVGEASICEDDMNGERTCTVTHTDLPSP 453
QY 511 IEKITSKAKGP-REPOVYTLPPSRDEL-T-KNOVSLTCLVGFPSDIAVWESNGQP- 566
DB 454 LKQITISPKGVALHPRDYLTPPAREQJLINESATITCLVGFPSADVFQVMQGPPLS 513
QY 567 ENNKKTTPPVLD--SDGSFFLYSKLTVDKSRMOQANVPSCVMEBALHNHTOKSLSP 624
DB 514 PEKVTATAPMEBPQAPGRVFAHSILITVSEEMNGERTTCVVAHEALPNRTEVTDKS- 572
QY 625 GLQLDFTCAEADGDELGLMTT 646
DB 573 -----TEGEVSADEGEFENLMAT 590

RESULT 44
QY 077600 PRELIMINARY; PRT; 86 AA.
AC 077600;
DT 01-NOV-1998 (Tremblrel. 08. Created)
DT 01-NOV-1998 (Tremblrel. 08. Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25. Last annotation update)
DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/LEU-3) (Fragment).
OS Lophocebus aetherius.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Lophocebus.
OX NCBI_TaxID=75566;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98320644; PubMed=9656488;
RA Harris E.B., Disocell T.R.;
RT "Nuclear gene trees and the phylogenetic relationships of the
RT mangabeys (primates: Papionini).";
RL Mol. Biol. Evol. 15:892-900(1998).
CC -1- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL
CC RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY
CC SIMILARITY).
CC -1- SUBUNIT: ASSOCIATES WITH P56-LCK (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC ONE V-LIKE AND 1 C2-LIKE DOMAIN.

```

```

DR EMBL; AF057390; AAC25134.1; -.
DR HSSP; P01730; 1CDY.
DR GO; GO:0016021; C:Integral to membrane; IEA.
DR InterPro; IPR007110; Ig-like.
DR Immunoglobulin domain; Glycoprotein; T-cell; MHC; Transmembrane.
FT NON_TER 1
FT DOMAIN <1 19 IG-LIKE V-TYPE DOMAIN.
FT DOMAIN 20 >86 IG-LIKE C2-TYPE DOMAIN.
FT CARBOHYD 79 79 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DISULFID 49 78 BY SIMILARITY.
FT NON_TER 86
SQ SEQUENCE 86 AA; 9476 MW; A9D97A88464FE93BE CRC64;

Query Match 11.2%; Score 382; DB 6; Length 86;
Best Local Similarity 85.9%; Pred. No. 5.5e-22;
Matches 73; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

QY 107 YICEVEDQKEVQLVFLGLTANSDTHLQGSLLTLESPPGSSPSVQCRSPRKNIOGK 166
DB 1 YICEVEDQKEVQLVFLGLTANSDTHLQGSLLTLESPPGSSPSVQCRSPRKNIOGK 60
QY 167 KTLVSQLELDGSGTWTCTVLQNGK 191
DB 61 RTLVSQLELDGSGTWTCTVNSQDOK 85

RESULT 45
QY 096AA6 PRELIMINARY; PRT; 618 AA.
AC 096AA6;
DT 01-DEC-2001 (Tremblrel. 19. Created)
DT 01-DEC-2001 (Tremblrel. 19. Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25. Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Lymph;
RA Strauberg R.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC017356; AAH17356.1; -.
DR PIR; S15590; S15590.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_V.
DR Pfam; PF00047; Ig_5.
DR SMART; SM00406; IGV_1.
DR PROSITE; PS50835; IG_LIKE; 5.
DR PROSITE; PS00290; IG_MHC; 3.
KW Hypothetical protein.
SQ SEQUENCE 618 AA; 67758 MW; 96DBD4C7C969E0A6 CRC64;

Query Match 11.2%; Score 382; DB 4; Length 618;
Best Local Similarity 21.8%; Pred. No. 1.1e-20;
Matches 149; Conservative 117; Mismatches 284; Indels 132; Gaps 25;

QY 8 RHLLVQLALLP-----AATQGNKVLGKGDVLTCTAQAOKKSIOFHMKNSNQIKI 61
DB 2 KHLMLFLLVADPRWVLSQVLOQMGAGLKPSETSLTCTGVGSGFSFGYWSWIRPPG 61
QY 62 LGNQ--GSFLTKGPSKLNDRADSRSL-WDQGNFPLIKNLKIEDSDT--YICEVEDQKE 116
DB 62 KGLEWIGEINHSSTWNPSTKSRVITISVDTSKQSLKLSVNAADTAVYYC----- 114
QY 117 EVQLVFLGLTANSDTHL--LQGSLLTLESPPGSSPS-----VQCRSPRKNIOGKCTL 169
DB 115 --ARVITRASPGRDGRGYMDVWGCGTIVTVSSGSASAPTLPLVSC----- 158
QY 170 SVSQLELDGSGTWTCTVLQNGKVEFKIDIVLAFQ-KASSIVKKEGEVFPPLAFT 228

```

Db 159 -----ENSPDSTSVAVGCLAQDPLPSITPSPMKYKNNSDISSTRG-----PSPVLK 205
 QY 229 VEKLTSSGGLMMQAEASSSSKMTITDLDKKEYSVKRVYQDPKLQMGKLLPLHL--TLFQ 286
 Db 206 GGGYAAATSOVLLPSSKVV-----MOGTDEHVVCVQHPNGNKKKXVPLFVLAELRP 255
 QY 287 ALPQVA-----GSGNLTALAEAKTGKHOEYNLVNMR-----ATOLQKNTCEVW 331
 Db 256 KYSVFPDPDPGFFGNPKRSKLTICQATGSPBROQVSWLREBKOVSGVTTDQVQAEAKS 315
 QY 332 GPTSPKMLSLKLENKAEVSKREKPEVWVLPNPMGMOCLSDSGVLLS-SNIXVLPW 390
 Db 316 GPFTYKVTSTLTIKESD-----WL---SQSNFTCRVDHRLGLTFQNNASSMVCVPQ 362
 QY 391 STFPVCPAPBPSSCDKTHTCPELLGGPSVFLPPPKKDTLMISRTREYTCVVVDVSHRP 450
 Db 363 DTAI-----RVFAIPPS-FASIFLTGSKTCLVLTDLTYD- 397
 QY 451 EYKFNMYVDGVEVHNATKPREEQYNSTYRVVSVLTVLHQDMLNGKREYKCKVSNKALPAP 510
 Db 398 SVTISSTRONGEAVKTHNINISHPNATTSVAGEASICEEDMNSGERFTCTVHTHDLPSR 457
 QY 511 IEKTISSKAGQP-REFQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEMESNQP-- 566
 Db 458 LKOTISRPKGVNLRDPDYLPPARQQLNRESATITCLVTGSPADVFQMMQGGQPLS 517
 QY 567 ENNYKTTTPVLD--SDGSFELYSLTVYDKSRWQGVFSGVNHKLNHNYTQKSLSP 624
 Db 518 PEKYVTSAPMPPEQAGRYFAHSILTVSEEWNTGETYTCVVAHEALPKRIVERTYDKS- 576
 QY 625 GLQDDETCAEADGELDGLMTT 646
 Db 577 -----TEGEVSADEEGFENLMAT 594

RESULT 46

Q8VCX7 PRELIMINARY; PRT; 613 AA.
 AC Q8VCX7;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Hypothetical protein.
 OS IGH-6.
 GN Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NC NCBITaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Salivary gland;
 RA Strausberg R.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC018315; AAH18315.1; -
 DR MGD: MGI:96448; IGH-6.
 DR InterPro: IPR007110; Ig-like.
 DR InterPro: IPR003006; Ig_MHC.
 DR InterPro: IPR003596; Ig_v.
 DR Pfam: PF00047; Ig; 5.
 DR SMART: SM00406; IGV; 1.
 DR PROSITE: PSS0835; IG_LIKE; 5.
 DR PROSITE: PSS00290; IG_MHC; 3.
 KW Hypothetical protein.
 SQ SEQUENCE 613 AA; 67855 MW; 41A9384DD4C22862 CRC64;

Query Match 11.1%; Score 380.5; DB 11; Length 613;

Best Local Similarity 22.5%; Pred. No. 1.4e-20;

Matches 157; Conservative 97; Mismatches 226; Indels 217; Gaps 28;
 QY 30 LGKKGDVELTCTAS-----QKSIQFHWKNSNQIKILGNQS--FLTKGPS 74
 Db 30 LMKPGASVKISKATGVTFSYWIEMVQKPGHGLEWIG---ELLPGSSSTVNEKFKG 85

QY 75 KLNRAD--SRSLMDQGNFLLIKLKIEDSDTYICEVEDQKEVOL-----VFGTLTA 127
 Db 86 KATFTADTSSNANVQ-----LSSLTSEDSAVYC-----ARLRGRWFDVWG--- 128
 QY 128 NSDTHLQGGSTLTLESF--FGSSPSVQCRSP-RGKNIQGGKTLVSVOLELDGSGTTC 184
 Db 129 -----AGTVTVSSESQSFPNVNPLVSCESPLSDKXILVAMGCLARPLPSTISFTWN- 180
 QY 185 TVLQNGKVEFKIDIVLAFQKASSIVYKKEGEQVFSPPLAFYBKLTGSELMMQAGR 244
 Db 181 -----YQNTETVQ----- 189
 QY 245 ASSSKWITFDLKNKEVVKRYTQDPKLQ-----GKKPLHLTLPLQALPQ 290
 Db 190 --GITFTPLRGGKYLATSOVLSFKGLBESDEYLVCKIHGGNRLHPIPI----- 242
 QY 291 YAGSNLTALAEAKTGKHOEYNLV-----MRATQLOKNTLCEVWGFTSPKMLSLKL 344
 Db 243 -----AVAMNPVNVVFPDPGSGAPRKSCLICEATNFTPPRITVSMWK 289
 QY 345 ENKEAKVSKREKPVWVN-----PEA-----GMQCLSDSGOVLLSEN 383
 Db 290 DKLVESEFTTDPVTIEKNGSTPQYKVISLTLSIEIDMLNINVTYCRVDHRLGTLFK-- 347
 QY 384 IKVLPWSTPVPCPAPBPSSCDKTHTCPELLGSPS---VFLPPPKKDTLMISRTREY 439
 Db 348 -----NVSTC---ASPSDTLITFTIIPSPAD-IFLSANIT 382
 QY 440 CVVVDVSHEDPEKENVYDGEVHNATKPREEQYNSTYRVVSVLTVLHQDMLNGKEYK 499
 Db 383 CLVSNLATYE-TLNTSMASQSGEPLETYKIMESHNPNGFSAKGVASVCVEBMMNRKEV 441
 QY 500 CYSNKAALPAIEKTISSKAGPRE-----POVYTLPPSRDELTKNQVSLTCLVKGF 553
 Db 442 CTVTRHDPSPQKFKISK---PNEVHKRPAPVYLPPARQQLNRESATITCLVAGESP 497
 QY 554 SDIAVEMESNQ--PENNYKTTTPVLD--SDGSFELYSLTVYDKSRWQGVFSGVNHK 609
 Db 498 ADISVQMLQRLGLQEKYVTSAPMPPEGAPPFYFTHSILTYTBEBWNSGERFTYCVVGH 557
 QY 610 ALHNHYTQKSLSPQLQDETCAEDGELDGLMTT 646
 Db 558 ALPHLVTERTYDKS-----TEGEVNAEEGFEENLMAT 589

RESULT 47

Q9BU10 PRELIMINARY; PRT; 597 AA.
 AC Q9BU10;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Hypothetical protein.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 NC NCBITaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lymph;
 RA Strausberg R.;
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC002963; AAH02963.1; -
 DR HSSP: P01825; 7FAB.
 DR InterPro: IPR007110; Ig-like.
 DR InterPro: IPR003006; Ig_MHC.
 DR InterPro: IPR003596; Ig_v.
 DR Pfam: PF00047; Ig; 5.
 DR SMART: SM00406; IGV; 1.
 DR PROSITE: PSS0835; IG_LIKE; 5.
 DR PROSITE: PSS00290; IG_MHC; 3.
 KW Hypothetical protein.
 SQ SEQUENCE 597 AA; 65274 MW; 2DAFA8FB7E055851 CRC64;

Query Match	11.1%	Score 377.5	DB 4	Length 597
Best Local Similarity	11.9%	Pred. No. 2.4e-20		
Matches 145	Conservative 115	Mismatches 274	Indels 127	Gaps 24
Qy	8	RHLLLVQLALLP-----AATQGNKVLGKKGDVLELTCTAASQKKSIFQFMKNSNQIKI	61	
Dy	2	KHLMFPELLVAAPRWLVLSQVQLQMGAGLLKPSLETSLTCGVYGGSPSGYVWSMIRPGP	61	
Qy	62	LGNG--GSFLTGKPSKLNDRADSRSL-WDQGNFLLIKNLKIEDSDT--YICEVEDQKE	116	
Dy	62	KGLMWIGINHSNGSTNNYPSLKSRTLTISVDPSKOLSLKLSVNAADRAVYVC-----	114	
Qy	117	EVQLVFLGELTNSDPHL---LQGGSLTLTLESPPSSPS-----VQCRAPRGNIQGGKTL	169	
Dy	115	--ARVITPASGTDGRKGMVDWGGSTITVYSSGASAPLTLPVYSC-----	158	
Qy	170	SVSOLELDSDGTWTCTVLQONKRYEKFIDIVLAFQ-RASSIVYKKEGEVFEPSPLAFT	228	
Dy	159	-----ENSPDSTSSVAVGLAQDFLPDSITFSMKYKNSDISSTRG-----PSPYLK	205	
Qy	229	VEKLTGSELMWQAEARASSKSMITFDLKAKESVSKAVTQDPKLOMGKLPRLH--TLPQ	286	
Dy	206	GGKVAATSTQVLLPSSKDV-----MQGIDENHYCKVQHPNGNKEKNVPLPIYALP	255	
Qy	287	ALPOYA-----GSGNLTLALBAKTGKLHGEVNLVWNR-----ATQLOKNLTCGEW	331	
Dy	256	KVSVPVPRPDGPFNGNPKRSKILCOATGSPROIQVSMIREGKQVSGVTTDVOVAEKES	315	
Qy	332	GPTSPKMLSLKLENKEAKVSKREKRPVWLNPEAGMOCLLSDSGQVLL-SNIKVLPT	390	
Dy	316	GPTTYKVTSTLTIKESD-----WL--SQGMFCRVADHRLGTFQNAASWCVDPDQ	362	
Qy	391	STPVPCPAPBEKSKCKHTTCBLLGSPVFLPFRPKOTLMSKRPETVGVVVDHSDP	450	
Dy	363	DTAL-----RVFAIPSP-FASIFLYKSTKLTCLVTLDTLTYD-	397	
Qy	451	EVEKNWYVDGVEVHNAKTPREEOYNSTYRVVSVTLVHODMLNKEKCKVSNKALPAP	510	
Dy	398	SVTISMTQNGEAVKTHINISESHNAIFSAVGEASICEEDMNGSERFTCVTHIDLSF	457	
Qy	511	IEKTISSKAKGP-REPOVYTLPPSHDELTKNVSLLGLVKGFPSPDIADVESNGDP--	566	
Dy	458	LKQITSRKGVALHHRPDVLLPAREQNLNRESATITCLVGTGSPADVFGVMQGGQPLS	517	
Qy	567	ENNYKTTPVLD--SDGSFPLYSKLTVDKSRQOQNVSCSYMEHALNNHTQKSLISP	624	
Dy	518	PEKYVTSAPMBEPQAPGRVFAHSILTVSEEMNTEGTYTQVVAHAPLNKRTERTVDKST	577	
Qy	625	G 625		
Dy	578	G 578		
RESULT 48				
Q9BOB8	ID	PRELIMINARY;	PRT;	597 AA.
Q9BOB8	Q9BOB8			
AC	Q9BOB8			
DT	01-JUN-2001	(Tremblrel. 17, Created)		
DT	01-JUN-2001	(Tremblrel. 17, Last sequence update)		
DT	01-OCT-2003	(Tremblrel. 25, Last annotation update)		
DE		Hypothetical protein.		
OS		Homo sapiens (Human).		
OC		Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
CC		Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.		
OX		NCBI_TaxID=9606;		
RN		SEQUENCE FROM N.A.		
RP		TISSUE=Muscle, and Lymph;		
RA		Strausberg R.;		
RL		Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.		
DR		EMBL; BC006180; AA006180.1; -		
DR		EMBL; BC001872; AA001872.1; -		

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NC NCB1_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lymph;
 RA Straubeberg R.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC019235; AAI19235.1; -
 DR InterPro; IPR007110; IG_Like.
 DR InterPro; IPR003006; IG_MHC.
 DR InterPro; IPR003596; IG_V.
 DR Pfam; PF00047; IG; 5.
 DR SMART; SM00406; IG; 1.
 DR PROSITE; PS00835; IG_Like; 5.
 DR PROSITE; PS00290; IG_MHC; 3.
 KW Hypothetical protein.
 SQ SEQUENCE 588 AA; 64438 MW; FC60DBAD2B39FD7 CRC64;
 Query Match 10.9%; Score 372.5; DB 4; Length 588;
 Best Local Similarity 21.9%; Pred. No. 5.6e-20;
 Matches 144; Conservative 115; Mismatches 273; Indels 127; Gaps 24;
 QY 8 RHLLVLQALALP-----AATGKNKVLGKGGDTVELTCTASQKKSIOFHMKNSNOIKI 61
 DB 2 KHLMPFLLLVAARRWLSQVQLQMGAGLIKSEITSLTGVYGGSFSGYMGWIRQPRG 61
 QY 62 LGNQ--GSFLTPKPSKLNDRADRSRL--WDQGNFPLIKNKLIEDSDT--YICEVEDQKE 116
 DB 62 KGLEWIGEINHSGSTVNPFLSKRSVTISVDTSKKQLSLKSVNADTAVYC----- 114
 QY 117 EVLLVFGLTANDTHL---LOGSLTLTLESPPGSSPS---VQCRSPKKNIOGSKTL 169
 DB 115 --ARVITRAPPGTDRGYMDVWGQTTVIVSSGSAAPLPLVSC----- 158
 QY 170 SVSQLELDGSGTWTCTVLQNKKEFKIDIVLAFQ--KASSIYKKEGEVEFPPLAFT 228
 DB 159 -----ENSPDSTSSAVAGCLAQDPLPDSITSMWKKNSDLSIRG-----FESVLR 205
 QY 229 VEKLTSGSELWQAEPRASSKSWITPDLKNEKVSVKRVTPQPKLQNGKKPLHL--TLRQ 286
 DB 206 GGRYAATSGVLRPSKDV-----MGTDEHVVCVKVQHPNGNKEKNVPLVIAELRP 255
 QY 287 ALPQYA-----GSGULTLALFAKTKLHQBVLNVMR-----ATQLQKNLTCFW 331
 DB 256 KSVFVPPRDPGFRGNPKSKLTCQATGFSRQIQVSWLRGKQVSGGVTTDQVQAEAKES 315
 QY 332 GPTSPKLMISLKLKENKAKVSKREKPPVWVLPAGMMQCLLSDSGQVLE--SNIKVLP 390
 DB 316 GPTTYKVTSLTLTKED-----WL---SQSMFTCVDRHGLTFQGNASSMCVPRQ 362
 QY 391 STFPVCPAPRPPKCDKTHTCPELLGGSVFLFPKPKKDTLMISRTBEVTCVVVDVSHEDP 450
 DB 363 DTAI-----RVFAIPPS--FASIFLTKSTLCTLVLDLTLYD-- 397
 QY 451 EYKFNMYVDGVEVNAKTKRREQVNSTYRVVSVLTVLHDWMLNGEKYKCKNSNKLAP 510
 DB 398 SVTISWTRONGEAVKTHNISESHPPATFSABASICEQDMNSGRFTCTVHTDLPSP 457
 QY 511 IEKTIKAKAGQ--REPOVYTLPPSRDELTKNGVSLTCLAVKGYPSDIAVEMESNGQ-- 566
 DB 458 LKQTTISRPKQVALHPRVYLLPPAREQLNRESATITCLVTGSPADVPQVMQROQPLS 517
 QY 567 ENNYKTTTPVLD--SDGSFFLYSKLTIVDKSRWQGNVFSVWHEALHNHTYQKSLSLG 623
 DB 518 PEKYVTSAPWPEQAPGRYPFAHSILTVSEEMNTGFEYTCVVAHEALPNRVTERTYDKS 576

DT 01-JAN-1999 (TrEMBLrel. 09, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE CD4 protein (Fragment).
 GN CD4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NC NCB1_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Neuroblastoma;
 RA Zverev V.V., Blinov V.M., Nedospasov S.A.;
 RT "Splice-mediated insertion of antisense and sense Alu repeats in human CD4 gene: identification of three exons of CD4 mRNA."
 RL Submitted (APR-1993) to the EMBL/GenBank/DBJ databases.
 DR EMBL; X81579; CA60883.1; -
 DR EMBL; S79267; AAB35273.1; -
 DR PIR; I60082; I60082.
 DR HSSP; P01730; 1CDY.
 DR GO; GO:004872; F:receptor activity; IEA.
 DR InterPro; IPR007110; IG_Like.
 DR PROSITE; PS00835; IG_Like; 1.
 KW Receptor.
 FT NON TER 71 71
 SQ SEQUENCE 71 AA; 7844 MW; A5C9DB4816135C66 CRC64;
 Query Match 10.5%; Score 357; DB 4; Length 71;
 Best Local Similarity 98.6%; Pred. No. 4.2e-20;
 Matches 70; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MNRGVPRHLLVLQALALPAAATGKNKVLGKGGDTVELTCTASQKKSIOFHMKNSNOIKI 60
 DB 1 MNRGVPRHLLVLQALALPAAATGKNKVLGKGGDTVELTCTASQKKSIOFHMKNSNOIKI 60
 QY 61 ILNGSGSFLTK 71
 DB 61 ILNGSGSFLTK 71
 RESULT 51
 ID 086TT1 PRELIMINARY; PRT; 375 AA.
 AC 086TT1;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Human full-length cDNA clone CS0DD006YL02 of neuroblastoma of Homo sapiens (Human).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NC NCB1_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Neuroblastoma;
 RA Genoscope;
 RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BX161420; CAD61894.1; -
 DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
 DR InterPro; IPR007110; IG_Like.

RL Dev. Dyn. 225:384-391 (2002).

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Spleen;

RA MEDLINE=2238257; PubMed=12477932;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkne R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,

RA Brownstein M.J., Uddin T.B., Tohyuki S., Carninci P., Prange C.,

RA Raha S.S., Longellano N.A., Peters G.J., Abramson R.D., Mallyah S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Villard S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

RA Villallon D., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahy J., Helton E., Kesteman M., Madan A., Rodriguez S., Sanchez A.,

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Rodriguez R.C., Greenwood J., Schmutz J., Myers R.M., Butterfield Y.S.,

RA Krzyzanski M.I., Skalska U., Smalins D.E., Scherch A., Schein J.E.,

RA Jones S.J., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length human

RT and mouse cDNA sequences.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).

RN [3]

RP SEQUENCE FROM N.A.

RC TISSUE=Spleen;

RA Klein S., Strausberg R.;

RL Submitted (Aug-2003) to the EMBL/GenBank/DBJ databases.

DR EMBL: BC056078; AAH56078.1; -.

KM Hypothetical protein.

SQ SEQUENCE 587 AA; 64933 MW; 646B5FD0FA14E8D CRC64;

Query Match 9.5%; Score 324.5; DB 13; Length 587;

Best Local Similarity 29.5%; Pred. No. 2.7e-16;

Matches 98; Conservative 54; Mismatches 115; Indels 65; Gaps 14;

QY 342 LKLEKKAAYSKREKV-----WLNPEAGMOCIL--SDSQVLL 380

DB 280 LKNGNOTBEGVVEVEDKRGYEATSYLSTRKEMDLD--LYSCVEHAESES-LQ 335

QY 381 ESKIKVLPWSTVPAPAPKSCDKHTPCBELLGGSVFLPPPKKDTLMTSRTEBVC 440

DB 336 EKMM-----SKSLMCDPTPTPTSIQVITP-----PSL-----ESTFEKKSATLTC 376

QY 441 VVVDVHEDEPEVFNMY-VDGVFNHAKTPREEQYNS--TYRVSVLTVLHODMNGKE 497

DB 377 LVSNMANSBDRISIMPKSGTQEIPLKTELGDATINDRNTYSVKGTTVCADENANDK- 435

QY 498 YKCKVSNKALPAIEKTISKAKQPREPOVYTLPPSRDELTKNQ-VSLTCLVKGFPSDI 556

DB 436 FVCKVEHTELSMKKEVFLFKKEGEYNTPSYVPPPELSEKRETTLTCLTVGFSPEL 495

QY 557 AYEM--ESNGQPENNYKTP-----PVLDSDGSPFLYSKLTVDKSRMOQGNVPSGVNH 608

DB 496 FVEMHNGEAVPKONTINTSINDELLPKGQSGKGFYLSLHTIDIDMDAGDSFCVGH 555

QY 609 EALNHVYTKSLSPG-----LQLDCTC 632

DB 556 ESLPLQLTQSRISDKSSGKPTNNVNSLVLSLDTCC 587

RESULT 54

Q27379 PRELIMINARY; PRT; 478 AA.

AC 072379

DT 01-OCT-2003 (TrEMBLrel. 25, Created)

DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Hypothetical protein DKFZp686K04218 (Fragment).

GN DKFZp686K04218.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=human rectum tumor;

RA Bloecher H., Boeher M., Mewes H.W., Weill B., Amid C., Oeanger A.,

RA Pobo G., Han M., Wiemann S.;

RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.

DR EMBL: BX538066; CAD97996.1; -.

KM Hypothetical protein.

FT NON TER 1

SQ SEQUENCE 478 AA; 51620 MW; 4AFCB541F3217CA1 CRC64;

Query Match 9.2%; Score 313.5; DB 4; Length 478;

Best Local Similarity 23.8%; Pred. No. 1.4e-15;

Matches 131; Conservative 65; Mismatches 177; Indels 177; Gaps 25;

QY 171 VSQLELDPSG-----TWCTVLQNKKEFKIDIVLAPQKASSIYKKEGEQVE 220

DB 17 LSQVQLQESGPELVKPSQTLSTCTV-----SGGSI----- 47

QY 221 FSPPLAFTVEKLTGSGELMWOAERASSSKSWITFDLKNKEVSKVVTQDPKLQMKLPL 280

DB 48 -----GSGDYFW-----SWI-----RQAP-----GRGLEW 67

QY 281 HLTLPQALPQYAGS-----GNITLALFAKGTGLHDEVN-----LYVNRATQLOKN 325

DB 68 -----MGYIYSGSYTYNPSLESRLSISIDSKNFSRLNSLTADTAIVFCAGVGIG 122

QY 326 LTCEVWG-----PTSPKLM-LSLKLEKKAAYSKR-----EKPVVNLNPEA 365

DB 123 TAFDWMGQGVTVTVSSASPTSPKVPPLSLDSTPQGNVNVALLVGFFQPELSTWMS 182

QY 366 GN-----WQCLSDSGOVLLSNIKVLPTWSTP-----VPCAP 399

DB 183 GGNVTAARNPPSQDASGLTYTSSQTLPATCQPGKSVTCVHKYTNPSQVTVPCVP 242

QY 400 EPKSCDKHTPCBELLGGSVFLPPPKKDTLMTSRTEBVC 459

DB 243 PPPPC-----C-----HRLSLHRLPALBD-LLLGEANLTCTLTGL-RASGATFTWPS 290

QY 460 GVEVFNHAKTPREEQYNS--TYRVSVLTVLHODMNGKEKCKVSNKALPAIEKTISKAK 519

DB 291 SGK--SANGPPEBRLCGYSVSVLPGCAQDMNGERTCTAAHPELKTPLANTTKS- 347

QY 520 GQPREPOVYTLPPSRDELTKNQ-VSLTCLVKGFPSDI-AVEMESNGQ--PENNYKTPPV 576

DB 348 GNTFPEPVHLLPPPEBELALNELVTLCLARGFSPPDVLVWLGSGQELPREKYLTMASR 407

QY 577 LD-SQG--SFLYSKLTVDKSRMOQGNVPSGVNHEALNHVYTKSLSPG----- 625

DB 408 QPSGQGTTFATYSILRAAEDWKGDFSCVSHALPLAFTQKTDRLAGKPTHVNS 467

QY 626 ---LQLDCTC 632

DB 468 VMAEYVDGTC 477

RESULT 55

Q29027 PRELIMINARY; PRT; 99 AA.

AC 029027

DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE CD4, allele 1 (Fragment).

OS Sus scrofa (Pig).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

OX NCBI_TaxID=9823;

RN [1]

RP SEQUENCE FROM N.A.

```
RC STRAIN=1183; TISSUE=Blood;
RA MEDLINE=93329116; PubMed=8335933;
RA Gueterson K.; Germana S.; Sundt T.M.; Sachs D.H.; LeGuern C.;
RT "Extensive allelic polymorphism in an exposed region of the
RT Malture. . . . .";
RL J. Immunol. 151:1365-1370(1993).
DR EMBL, X65629; CAA46583.1; -.
DR PIR, I47131; S21461.
DR HSSP, P01730; 1CDY.
DR InterPro: IPR007110; IG_1like.
DR InterPro: IPR003596; IG_v.
DR Pfam: PF00047; IG_1.
DR SMART, SM00406; IGv_1.
DR PROSITE, PS00835; IG_LIKE; 1.
FT NON_TER 1
FT NON_TER 99
SO SEQUENCE 99 AA; 11170 MW; 40BF080699CF5D0C CRC64;

Query Match 8.9%; Score 305.5; DB 6; Length 99;
Best Local Similarity 60.2%; Pred. No. 6,1e-16;
Matches 59; Conservative 16; Mismatches 22; Indels 1; Gaps 1;

Qy 32 KKGDTVELTCTASQKSIQFHMKNSTQIKINGQSF--TKGPSKLDNRADRSRLNDQG 90
Db 1 KAGDLAELPCHSSSQKKNLPFWKNSTQIKLGGHGFMTASVTELSRLDSKKNMMDHG 60

Qy 91 NFPLIKLKIEDSDTYICEVEDQKEEYQLVFGLTAN 128
Db 61 SFPLIKLEVTDSGIYCEVEDKRIEYQLLVFLITAS 98

RESULT 56
Q9NPP6 PRELIMINARY; PRT; 416 AA.
AC Q9NPP6:
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Immunoglobulin heavy chain variant (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN 1
RP SEQUENCE FROM N.A.
RA Pubmed R. Estivill X., Escarceller M., Sunoy L.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RN 12
RP SEQUENCE FROM N.A.
RA Auffray C., Anorge W., Ballabio A., Estivill X., Gibson K.,
RA Lebrach H., Pouetka A., Lundberg J.;
RT "The European IMAGE consortium for integrated molecular analysis of
RT human gene transcripts.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL, AL389978; CAB97534.1; -.
DR HSSP, P01789; IMCP.
DR InterPro: IPR007110; IG_1like.
DR InterPro: IPR003006; IG_LIKE.
DR InterPro: IPR003596; IG_v.
DR Pfam, PF00047; IGv_1.
DR SMART, SM00406; IGv_1.
DR PROSITE, PS00835; IG_LIKE; 3.
DR PROSITE, PS00290; IG_MHC; 2.
FT NON_TER 1
FT NON_TER 99
SO SEQUENCE 416 AA; 44786 MW; 8C41708BB8AB4687 CRC64;

Query Match 8.8%; Score 302; DB 4; Length 416;
Best Local Similarity 27.8%; Pred. No. 8,9e-15;
Matches 102; Conservative 47; Mismatches 140; Indels 78; Gaps 16;

Qy 329 EWNG-----PTSPKLM-LSLKLKNEKAVSKR-----EKPVVNLNPEAGM- 367
Db 64 DVMGGTTTVVSASPTSPKFLPLSLDSTPDGNNVVVACLVGFFPQEPISLVMSSESGN 123
```

```
Qy 368 -----WCLLSDSGQVLLSINIKVLPWSTP-----VPCPADEPK 402
Db 124 VTANFPSPQASDLDVTTSSQLPLPNTQCPDGKSVCHVGHYTNPSQDVPVPCVPPR 183

Qy 403 SCDTHTCPELGGPSVFLPPPKXTLMTSRPEVTCVVVDVSHEDPEVKENYVDGVE 462
Db 184 PC-----C-----HPRLSLHRPALED--LLIGSEANLCTLTGL--RDASGATFTWPSGK 231

Qy 463 VHNAKTPREQVNSTRVVSVLTVLHQDMNKGVEYCKSNKALPAPIETISKAQGP 522
Db 232 --SAVQAPPERDLCCYSVSSVLPFGCAQPMNHGETFTCTAAHPELKTPTLANITKS--GNT 288

Qy 523 REPQVYTLPPSRDELTKNQ--VSLTCLVKGFYPSDIAVEMESNQ--PENNYKTPPYLD- 578
Db 289 FRPVPVHLPPSEELANLVLTLCLARGSPKDVLRVNIQSGSELPREKILYTWASQEP 348

Qy 579 SDG--SFFLYSKLTVDKSRWQGNVFSQVMEHLENHNYTQKSLSPG----- 625
Db 349 SQGTTTAVVSIIRVAEDMKKGDPTFCMGHEALPLAFTQKTIIDRLAKPTHVNVSVM 408

Qy 626 LQIDETC 632
Db 409 AEVDGTC 415

RESULT 57
Q96K68 PRELIMINARY; PRT; 494 AA.
AC Q96K68:
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein FLJ14473.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN 1
RP SEQUENCE FROM N.A.
RC TISSUE=Mammary gland;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Magatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishi S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuho Y.,
RA Niinomiya K., Iwayanagi T.;
RT "NDO human cDNA sequencing project.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL, AK027379; BAB55072.1; -.
DR PIR, S21205; S21205.
DR InterPro: IPR007110; IG_1like.
DR InterPro: IPR003006; IG_LIKE.
DR InterPro: IPR003596; IG_v.
DR Pfam, PF00047; IGv_1.
DR SMART, SM00406; IGv_1.
DR PROSITE, PS00835; IG_LIKE; 4.
DR PROSITE, PS00290; IG_MHC; 1.
KM Hypothetical protein.
SO SEQUENCE 494 AA; 53088 MW; 9A1D7A85A8E4C0E CRC64;

Query Match 8.8%; Score 300.5; DB 4; Length 494;
Best Local Similarity 21.8%; Pred. No. 1,5e-14;
Matches 139; Conservative 82; Mismatches 208; Indels 209; Gaps 27;

Qy 30 LGKKGDTVELTCTAS--QKSIQFHW-----KNSQIKLNGQSF-----TKGPSKL 76
Db 30 LVPPGSLRISCAASGLSFTYAMNWRQAPGGLVWVSSISRSRDYIYRDSVKGRFTI 89

Qy 77 NDRADRSRLNDQGNFPLIKLKIEDSDTYICEVEDQKEEYQLVFGLTANSDTHLQG 136
Db 90 S-RDNKNSLYLQMN-----SLRVDDTAVYICARDCNGAI--CYGFSF-----WG 132
```

```

QY 137 QSLTLTLESPSSPEVQCRSPKRNIGCGKTLISVQLDQSGTCTTCLVLOKQKVERK 196
DB 133 QGLTLTVSSASPSPT-----KYFPLSLCSTDPG-----NVV 164
QY 197 IDIVLAF--OKASSIVYKKEGQVEF-SFPLAFTEKLTGSGELMWQABRASSKSWIT 253
DB 165 IACLVOGFPPQPELSTWSESGQGVARNP-----PSODASGDLT-----TTSO---- 210
QY 254 FDLKNKEVSVKRYTQDPKLOMGKCLPLHLTLPOALPOYASGNLTALBAKTGKLGQEVN 313
DB 211 -----LTLF----- 214
QY 314 LVVNRATQ--LQKNLTCEVWGPTSPKMLSLKLENKAKYSKRKPPVWVLPNPAQMWQCL 371
DB 215 -----ATQCLAGKSVCVCHVHYTNP----- 234
QY 372 LQDSGOVLESNIKVLPTWSTPVPAPAPPKSCDKHTHCEBLLGSPVFLPPKPDITLM 431
DB 235 ---SQDVTVPQCPPTPTPTSPPTPTSPSCCH-----PRLSLHRPALBD-LI 279
QY 432 ISRTPEVTCVVVDVSHEDPEVKFNMYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLAHOD 491
DB 280 LGSEANLTCTLTGL-RDASGVTFWTTPSSGK--SAVOGPPERDLGCGYSVSVLPQCAEP 336
QY 492 WLNGKEYKCKVSKNALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQ-VSLTCLVKG 550
DB 337 MNHGKFTCTAAYPESKTPPLTATLSKS-GNTRPREVHLPPPSSEELALNELVTLTCLAG 395
QY 551 FVPSDIAVEMESNGQ--PENNYKTTTPVL-D-SG--SFPLYSKLTVDKSRMQGNVFSQS 605
DB 396 FSKPDVLVWMLQSGQLPREKYLTMASROEPPSGITTFATVTSILRVAEDMKKGDITFSCH 455
QY 606 VMHEALNHNYTKSKLSLSPG-----LQLDFTC 632
DB 456 VGHFALPLAFTQKTIIDRLAGKPTHVNVSVVMAEVDGTC 493

```

```

RESULT 58
Q96KX8
ID 096KX8 PRELIMINARY; PRT; 496 AA.
AC 096KX8;
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RA Strausberg R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC016369; AAH16369.1; -
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; Ig_4.
DR SMART; SM00406; IgV_1.
DR PROSITE; PS00835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 1.
KW Hypothetical protein.
SQ SEQUENCE 496 AA; 53391 MW; D346929849040D69 CRC64;

```

```

Query Match 8.7%; Score 298; DB 4; Length 496;
Best Local Similarity 23.3%; Pred. No. 2.3e-14;
Matches 132; Conservative 69; Mismatches 209; Indels 156; Gaps 21;
QY 124 GLTANSDPTHLQSGTLTLTLESPGSSPVQ---CRSPKRNIGCGKTLISVQL-----E 175
DB 29 GLVKSSET-----LSLTCTVSGGSISSSYWGMIRQPPKGLMIANTYTSGLITTYNPS 83

```

```

QY 176 LQDSGTWCTVLQONQKVEFKIDIVLAFQKASSIVYKKEGQVEFSFPLAFTEKLTGS 235
DB 84 LKSRVTLISDTSKQ-----LSLKVRSVTADTAVY-----FCARHGSRSRGRTA 129
QY 236 GELMWQABRASSKSWITFDLKNKEVSVKRYTQDPKLOMGKCLPLHLTLPOALPOYASG 295
DB 130 IDYWGQGLTVVSSASPT-----SKRV-----FPLSLCSTP-----DG 163
QY 236 NITLAL-----EAKTKLHQBVLVNRATQ--LQ 323
DB 164 NVVIACLVOGFPPQPELSTWSESGQGVARNPPEPSODASGDLTYTSSQTLTPATQCLAG 223
QY 324 KVLTEVWGPTSPKMLSLKLENKAKYSKRKPPVWVLPNPAQMWQCLSDSGVLESN 303
DB 224 KSVTCHVHYTNP-----SQDVTVPQCP 245
QY 384 IKVLPTWSTPVPAPAPPKSCDKHTHCEBLLGSPVFLPPKPDITMISRTPEVTCVV 443
DB 246 VSTPTPTSPPTPTPTSPSCCH-----PRLSLHRPALBD-LLGSEANLTCTLT 293
QY 444 DVSHEDPEVKFNMYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLAHQVNLNGKEYCKYS 503
DB 294 GL-RDASGVTFWTTPSSGK--SAVOGPPERDLGCGYSVSVLPQCAEPWNHGKTFCTAA 350
QY 504 NKALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQ-VSLTCLVKGFPEDIAVEMES 562
DB 351 YPESKTPPLTATLSKS-GNTRPREVHLPPPSSEELALNELVTLTCLAGSPQDVLFRWLQ 409
QY 563 NGQ--PENNYKTTTPVL-D-SG--SFPLYSKLTVDKSRMQGNVFSQSVMHEALNHNYTQ 617
DB 410 GSGELPREKYLTMASROEPPSGITTFATVTSILRVAEDMKKGDITSCWGHFALPLAFTQ 469
QY 618 KSLSLSPG-----LQLDFTC 632
DB 470 KTIIDRLAGKPTHVNVSVVMAEVDGTC 495

```

```

RESULT 59
Q72374
ID 072374 PRELIMINARY; PRT; 492 AA.
AC 072374;
DT 01-OCT-2003 (TREMblrel. 25, Created)
DT 01-OCT-2003 (TREMblrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)
DE Hypothetical protein DKFZp686C02218 (Fragment).
GN DKFZp686C02218.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Human rectum tumor;
RA Bloeker H., Boecher M., Mewes H.W., Weil B., Amlid C., Osanger A.,
RA Robo G., Han M., Wiemann S., EMBL/GenBank/DBJ databases.
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BX538077; CAD98001.1; -
KW Hypothetical protein.
FT NON TER
SQ SEQUENCE 492 AA; 53776 MW; 1E7A15760F0CA74B CRC64;

```

```

Query Match 8.7%; Score 297; DB 4; Length 492;
Best Local Similarity 27.7%; Pred. No. 2.7e-14;
Matches 101; Conservative 46; Mismatches 140; Indels 78; Gaps 16;
QY 331 NG-----PSPKLM-LSLKLENKAKYSK-----EKPPVWVLPNPAQMWQCLSDSGVLESN 367
DB 142 WQSGTLTVVSSASPSPTSPKPLSLDSTPDQGNVAVACLVQGFPPQPELSTWSESGQNT 201
QY 368 ---WQCLSDSGOVLESNIKVLPTWSTP-----VPCAPAPPKSC 404
DB 202 AANFPQSASGDLTYTSSQTLTPATQCPDGKSVTCHVHYTNPQSDVYTPCVPVPPPPC 261

```

```
Qy 405 DKHTCEBLLGSPVFLFPKPKDTLMISRTPEVTCVVDVSHEDPEVKFNMVYDGEVH 464
Db 262 -----C-----HRLSLHRLPALD-LILGSEANLCTLTGL-RASGATFTWTSBGK-- 307
Qy 465 NAKTPREBQYNSYRVVSVLTVLHQLMNGEKYCKVSNKALPAPIEKTISKAGQPRE 524
Db 308 SAVGQPEPRDLGGCVSVSVLPGCAQPPWTHGETFTCTAAHDELKTPLTANITKS-GNTFR 366
Qy 525 POYVTLPPSRDELTKNQ-VSLTCLVKGPYFSDIAVESNQ--PENNYKTPPLVD-SD 580
Db 367 PEVHLPPSEELALNELVTLTCLARGSPKDVLRVLQSGQELPREKYLTMASRQEPSQ 426
Qy 581 G--SFFLYSKLTVDSRMQGNVSCSVMEALHNHYTOKSLSPG-----LQ 627
Db 427 GTTFATYSLIRVAEDMKKQDTSQVGHFLPLAFTOKTIDRLAGKPTHVNVSVMAE 486
Qy 628 LDFTC 632
Db 487 VDGTC 491
```

RESULT 60

```
Q8NSK4 PRELIMINARY; PRT; 499 AA.
AC O8NSK4;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Blood;
RA Straubeberg R.;
RL EMBL; BC032249; AAH32249.1; -.
DR InterPro; IPR003599; IG.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003597; IG-cl.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; IG; 4.
DR SMART; SM00409; IG; 4.
DR SMART; SM00407; IGH1; 2.
DR PROSITE; PSS0835; IG_LIKE; 4.
DR PROSITE; PSS0290; IG_MHC; 1.
KW Hypothetical protein.
SQ SEQUENCE 499 AA; 53376 MW; 93A5C89582054F32 CRC64;
```

Query Match 8.7%; Score 296; DB 4; Length 499;

Best Local Similarity 25.2%; Pred. No. 3.3e-14; Indels 128; Gaps 24;

```
Matches 131; Conservative 60; Mismatches 201;
Qy 219 VESFPLAFTVEKLTG-----SGELMQAERASSKSWTFDLAKKEV 261
Db 1 MERGLSVPLVAILKGVOCEQLVESGGGVRRPGSL-----RLSCATSGFTD---DS 51
Qy 262 SVKRVYDDPKLQMGKLP--LHLTLPOALPOYAGS--GNLTALAEATKGLHDEVNLVVM 317
Db 52 GASVVRQAP-----GKGLEWVSSIMNGSGSTVADSVGRFTISRDNKNSLYIQMNSLRV 107
Qy 318 RATOLQ-----KNLTC-----EVMG-----PNSPKLM--LSLKLKNEAK 350
Db 108 EDTALYYCARDPTYYCGSGCLGYVMVWGKTTVTVSASPTSPKVPFLSLCSTQPDGN 167
Qy 351 VSKR-----EKPVVNLPEAGM-----WQCLLD 374
Db 168 VVLAICLVGFPFGPRLSVTWSSESGGVYANRFPSPQASGDLVYTTSSQLTLPATQCLAGK 227
```

```
Qy 375 SGQVLESNIK--VLPTWSTPVPQAPPEKSCDKHTCPQLG---PSVFLFPKPKDT 429
Db 228 S-----YCHVAKHTNNSQDVTVPDPSTPTSPSPSTPPSPSCCHPRSLHRLPALD- 282
Qy 430 LMSRTPPEVTCVVDVSHEDPEVKFNMVYDGEVHNAKTPREBQYNSYRVVSVLTVLH 489
Db 283 LILGSEANLCTLTGL-RDASGVTFPTWTSBGK--SAVGQPPRDLGGCVSVSVLPGCA 339
Qy 490 QDMLNGEKYCKVSNKALPAPIEKTISKAGQPREQVYTLPPSRDELTKNQ-VSLTCLV 548
Db 340 EPMNHGKTFCTAAYESKTPLTATLTKS-GNTFRPEVHLPPSEELALNELVTLTCLA 398
Qy 549 KGFYPSDIAYEWSNQ--PENNYKTPPLVD-SDG--SFFLYSKLTVDSRMQGNVFS 603
Db 399 KGFSPKDVLRVLQSGQELPREKYLTMASRQEPSQGTTFATYSLIRVAEDMKKQDTS 458
Qy 604 CSVMHEALHNHYTOKSLSPG-----LQLDFTC 632
Db 459 CMVGHFLPLAFTOKTIDRLAGKPTHVNVSVMAEVDGTC 498
```

RESULT 61

```
Q8NCL6 PRELIMINARY; PRT; 493 AA.
AC Q8NCL6;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein FLJ90170.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Mammary gland;
RA Isegai T., Oca T., Nishikawa T., Hayaishi K., Otsuki T., Sugiyama T.,
RA Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hiro Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Kojima S., Nagahari K.,
RA Masuno Y., Ono T., Okano K., Yoshikawa Y., Aotsuka S., Sasaki N.,
RA Hattori A., Okumura K., Iwayanagi T., Ninomiya K.,
RT "NEDO human cDNA sequencing project.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK074651; BAC1114.1; -.
DR InterPro; IPR003599; IG.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003597; IG-cl.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; IG; 4.
DR SMART; SM00409; IG; 4.
DR SMART; SM00407; IGH1; 2.
DR PROSITE; PSS0835; IG_LIKE; 4.
DR PROSITE; PSS0290; IG_MHC; 1.
KW Hypothetical protein.
SQ SEQUENCE 493 AA; 53224 MW; 12ECD7E094777101 CRC64;
```

Query Match 8.6%; Score 294; DB 4; Length 493;

Best Local Similarity 22.9%; Pred. No. 4.7e-14; Indels 212; Gaps 30;

```
Matches 147; Conservative 77; Mismatches 207;
Qy 25 GNRVVLGKKGDTVELTCTAS--OKKSIQFHWKNSNO-----IKILNQG-----SFLTK 71
Db 27 GGGVVL--PGSGRLSCAASGFRFRDYDMHWNRQSGEGLEWALLWYGTCTKYVSDSVK 84
Qy 72 GPKSLNDRADSRSLMDQGNFPIILNKLTIEDSDTYICEVEDQKEEVQLVFGLTANSBT 131
Db 85 GRLTVS-RDNYKNLTLY-----LEMKSIGEDTAAYYC-ARDQ-----GYAGYGVF-----D 128
Qy 132 HILQGSLLTLLESPPSGSSPSVQGRPRGNIGCGTTLSVSQLDQDSGTWCTCTVLQNRK 191
```

[illegible]

	RESULT	62
Q9BRV0	PRELIMINARY;	PRT; 500 AA.
ID Q9BRV0		
AC Q9BRV0		
DT 01-JUN-2001	(TREMBLrel. 17, Created)	
DT 01-JUN-2001	(TREMBLrel. 17, Last sequence update)	
DT 01-OCT-2003	(TREMBLrel. 25, Last annotation update)	
DE	Hypothetical protein.	
OS	Homo sapiens (Human).	
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.		
OX NCBI_TaxID=9606;		
RN [1]		
RP SEQUENCE FROM N.A.		
RC TISSUE=Prostate;		
RA Strauberg R.;		
RL Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.		
DR EMBL; BC005951; AAH05951.1; -.		
DR HSSP; P01789; IMCP.		
DR InterPro: IPR007110; IG-like.		
DR InterPro: IPR003006; IG_MHC.		
DR InterPro: IPR003596; IG_V.		
DR Pfam; PF00047; Ig; 4.		
DR SMART; SM00406; IgV; 1.		
DR PROSITE; PSS0835; IG_LIKE; 4.		
DR PROSITE; PSS0290; IG_MHC; 1.		
KM Hypothetical protein.		
SQ SEQUENCE 500 AA; 54154 MW; 0A9BF43F2A3CC6D9 CRC64;		
Query Match	8.5%; Score 289; DB 4; Length 500;	
Best Local Similarity	22.6%; Pred. No. 1.2e-13;	
Matches 136; Conservative 67; Mismatches 203; Indels 196; Gaps 26		
Oy 118 VOLLVGLT-ANDSDTLLOGSGSLTLTPSPSSPVYCRSPRKNTGGKTLTVSQTLL	176	
Dd 7 ILFLVAATAAGASOVLHVDSGAELWS---PEASVRVSCKT-----	43	
Oy 177 QDSGTWTCTVLQNQCKVEERKIIVLAIFQKASSIYYKKKEGVQVERSPFLATVEKLTGSG	216	

```

Db      44  --SG-----YAFHTTSTII-----WVOAPQOG 63
Qy      237 ELWMOERASSKSWITFDLKNKEVSVKRVOTDPRLOMGKTLPLHLTLPOALPOYAGSGN 296
Db      64  LEN-----MGWISPSDDN-----TRAKKF-----QGR 86
Qy      297 LTLALNAKQKGLHOEV-----NLVWRATLOQNLTC-----EWVG-----
Db      87  VTLTTTSTSTVYMWELRSLRSDDTAVVYCARRYCSYSSCONDYYYYYMDVMGKGTTVVS 146
Qy      333 ---PTSPLM--LSIKLENKEAKYSKR-----EKPMVLNPEAGM-----367
Db      147 SASPTSPKPKPLSLCSTGTPDGNVVAICLVQGFPEPRLSVTMSBEGOGVTARNPSPDA 200
Qy      368 -----WQCLLSDSGOVLLESNIK--VLEPTWSTPVPCPAPRPSCKDTHTC 410
Db      207 SGDLTYTTSSQTLPLATQCLAGKS-----VTCHVKHVTNSQDUTVPCBPVSPPTPSPSTP 262
Qy      411 PELLGG---PSVLEPFPKPKDOLMTLSRIPEVTCVQVVDVSHDEPEKFMWYVDGVEYNNAK 466
Db      263 PTPSPSCPPRSLSHRPALED--LLLGSEANLTCVTLGI--RDSGVTFTWTPSSG--SAV 311
Qy      468 TKREEOYNSTYRVVSVLTVLHQDWLNGKEYKCKSNALPAPLEKTIKSAKGQPREBOV 527
Db      319 QGPRPDRDLGGCVSVSVLSGCAEPNMNHGKTFCTAAYPEKSTPLATLSTKS--GMTFPREV 377
Qy      528 YTPSPRDELTKNQ--VSLTCLVKGYPSPDIAMWMSNQ--PENNYKTPPVLD--SDG--588
Db      378 HLLPPESEIALNEVLTITLCIARGSPKDVLRWLOQSOQLPREKYLTWASROEPSGTT 437
Qy      582 SFELYEKLTVYDSRMOOGVVFCSVMHEALHNHYTQKSLSLSPG-----LQDE 630
Db      438 TPAVYIILVAADMKGDPTSCMGHEALPLAFQETIIDLAKGPTHVNVSVYMAEVDG 497
Qy      631 TC 632
Db      498 TC 499

```

RESULT	63
Q96DK0	
ID	Q96DK0
PRELIMINARY;	PRT; 496 AA.
AC	Q96DK0;
DT	01-DEC-2001 (TrEMBLrel. 19, Created)
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE	Hypothetical protein FL025298.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX	NCBI_TaxId=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=Gastric mucosa;
RA	Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,
RA	Hochu T., Hirooka S., Murakawa K., Takiguchi S., Kusano J.,
RA	Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,
RA	Sugiyama Y., Hata H., Nakagawa K., Mizuno S., Morinaga M., Kawamura M.,
RA	Sugiyama T., Irie R., Otsuki T., Sato H., Nishikawa T., Sugiyama A.,
RA	Kawakami B., Nagai K., Isogai T., Sugano S.
RT	"NEDD human cDNA sequencing project."
RL	Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.
DR	EMBL; AK058027; BAB71633.1; ..
DR	InterPro; IPR007110; IG_1like.
DR	InterPro; IPR003006; IG_MHC.
DR	InterPro; IPR003596; IG_V.
DR	Pfam; PF00047; IG_4.
DR	SMART; SM00405; IG_1.
DR	PROSITE; PS50835; IG_LIKE; 4.
DR	PROSITE; PS00290; IG_MHC; 1.
KW	Hydrophobic protein.
Q0	SEQUENCE 496 AA; 53533 MW; C72EE1E247C86FED CRC64;

Query Match 8.5%; Score 288.5; DB 4; Length 496;
 Best Local Similarity 22.7%; Pred. No. 1.2e-13;
 Matches 134; Conservative 65; Mismatches 192; Indels 199; Gaps 26;

129 SDTHLLQ-GGSLTLTLESPSSPSVQCRSPRGKNIQGGKTLVSQLELDGSGTWCTVL 187
 19 SQVHLVOSGAEIKM-----PGSSVYVSC----- 41
 188 QNKKVEFKIDIVLAFOKASSIYKKEGEVFSPLAFT-VEKLTGSGELMMQARAS 246
 42 -----KASANNFRS-----YAFTWBQAPQGLQW----- 66
 247 SSKSMITFDLKNKEVSVKRVTDPRKLMGKKLPLHLTPQALPOYAGS--GNLTALAEAK 304
 67 -----MGGLIPNF-----GAPYNAQNFQORVITISADS 94
 305 TGKHLQEVNLVVMRATOLQ--KNLT-----CEVWG-----PTSPKLM-L 340
 95 TTYVYMETLSLTFEDTAFYVYCGRGLTYGSGSYLLQHMGGTLVTVSASPTSPKVPPL 154
 341 SLXENKEAKVSKR-----EKPVWVLANPEAGM----- 367
 155 SLCTQPDGNVNIACLVGGFPPOEPLSVTWSESGQVTAARNPPSQASGDLTYTSSQLT 214
 368 ---WQCLSDSGQVLLLESNIK--VLPTWSTVPQCPAPEPKSCDKHTCPPELLGG---PSV 419
 215 LPAQOCLAGKS-----VTCHVGHYNTPSODVTPPCVPSTPTPTSPPTPTSPBCCHPRL 270
 420 FLPPPKPDTLMISRTPEVTCVVVDVSHEDPEVKFNNVYDVEVNAKTKRREQDYNSTY 479
 271 SLHPRALSD-LLLSEANLTCTLTGL-RDASGVTFMTWTPSSGK--SAVQGPDRDLGGCY 326
 480 RVSVVLTVLHDDMLNGEKYKCKVSNKALPAPIETKISAKQPREPOVYITLPSRDELTK 539
 327 SVSSVLPGCAFPNMHGKFTCTTAAYPESKPTLTATLSNS--GNTRPEVHLPPSEELAL 385
 540 NQ--SLTCLVKGFPSPDIAVEWESNGQ--PENNYKTPPVLD--SDG--SFLYSLKLTVDK 593
 386 NELVTLCIARGFPPKVLVAMLOGSDLPREKLTLMASROBPQGGTTTFAVISILVAA 445
 594 SRMOQGNVFGSVNHEALHNHYTKSLSPG-----QLDDETQ 632
 446 EDWKKGDTFSCVHGHEALPLAFTQKTIDRLAGKPTHVNVSVMAEVDGTC 495

Db

RESULT 64

Q91WR1 PRELIMINARY; PRT; 488 AA.

AC Q91WR1; Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Hypothetical protein.
 GN IGH-VJ558 OR AI893585.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Strauberg R.;
 RC Tissue=Kidney;
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC013539; AAH13539.1; -
 DR MGD; MGI:96486; Igh-VJ558.
 DR InterPro; IPR007110; Igh-like.
 DR InterPro; IPR003006; Igh_MHC.
 DR InterPro; IPR003596; Igh_V.
 DR Pfam; PF00047; Igh; 4.
 DR SMART; SM00406; IGV; 1.
 DR PROSITE; PS50835; IGH LIKE; 4.
 DR PROSITE; PS00290; IGH_MHC; 2.

KM Hypothetical protein.
 SQ SEQUENCE 488 AA; 52964 MW; F12068460B400B9D CRC64;

Query Match 8.4%; Score 288; DB 11; Length 488;
 Best Local Similarity 23.8%; Pred. No. 1.3e-13;
 Matches 128; Conservative 72; Mismatches 179; Indels 158; Gaps 29;

147 PGSSPSVQCRSPRGKNI-----QGGKTLVSQLELDGSGTWCTVLQNKVYEF 195
 33 PGASVLSKCA-SGYTTTDVYVMMVQSHKSLKEWICDIPNYGCT-----SYNQK---F 83
 196 KIDIVLAFOKASSIYKKEGEVFSPLAFTVEKLTGSGELMMQARASSSKWITFD 255
 84 K-GKATLVKSSSIAYWQ-----LNNLTSDSAVYVYCARGPVYYSYPSYD 128
 256 LKN--KEVSVKRYTQDPKLMGKKLPLH-LTLQALP-----QYASGNLTAL 301
 129 RGYWGGTLVTVSAPD---AREPTIYPLTFQALSSDVIITGCLIHDFPSTGM--- 180
 302 EAKTGKHLQEVNLVVMRATOLQKNLTCEVWGPTSPKLMSLXENKEAKVSKKEKPVWL 361
 181 -----NVT---WG-----KSGKIDITV 194
 362 N-PEAGMWQCLSDSGQVLLLESNIKVLPTWSTVPQCPAPEPKSCDKHT-----TC 410
 195 NFPRA-----LASGRYTWSQL-TLPA-----VECEGSGSVCSVGHSDNPQELNVNC 243
 411 PELIGSGSVLPFPKPK-----DTLMISRTPEVTCVVVDVSHEDPE-VKFNWY 457
 244 FGICSPPTT---PPPSQPSLSLQRPALDILLGSDASTICTLNLG--RDPEGAVPTW- 297
 458 VDGVEVNAKTKRREQDYNST---YRVSVLTVLHDDMLNGEKYKCKVSNKALPAPIETK 514
 298 -----EPSTGDAVQKAVQNSCCYSVSVLPQCAEWNWSGASFCTVTHPESDT-LTGT 352
 515 ISRAKQPREPOVYITLPSRDELTKNQ-VSLTCLVKGFPSPDIAVEWESNGQ--PENNY 570
 353 IAVTVNTPFPQVHLPPREBELALNELVSLTCLVAFNKEVLRKHLHNEELSPSYL 412
 571 KTTPEVLDSG--SFLYSLKLTVDKSRMOQGNVFGSVNHEALHNHYTKSLSPG 625
 413 VFEPLEKPEGATTVLTVSVLRVSAEIMKGGDQYSQVHGHEALPMNFTQKTIDRLSG 469

Db

RESULT 65

Q9UP60 PRELIMINARY; PRT; 384 AA.

AC Q9UP60; Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE SNC73 protein.
 GN Homo sapiens (Human).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Zheng S., Cao J., Cao W., Cai X., Gang L.;
 RT "Identification and characterization of SNC73, a gene which is down-regulated in colorectal cancer."
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF067420; AAC19365.1; -
 DR HSSP; P01825; 7FAB.
 DR InterPro; IPR007110; Igh-like.
 DR InterPro; IPR003597; Igh_C1.
 DR InterPro; IPR003006; Igh_MHC.
 DR Pfam; PF00047; Igh; 3.
 DR SMART; SM00407; IGH1; 2.
 DR PROSITE; PS50835; IGH LIKE; 3.
 DR PROSITE; PS00290; IGH_MHC; 1.
 SQ SEQUENCE 384 AA; 40947 MW; BA7AD3CA5A9DD48 CRC64;

Query Match 8.3%; Score 282.5; DB 4; Length 384;
Best Local Similarity 26.5%; Pred. No. 2.5e-13;
Matches 92; Conservative 46; Mismatches 136; Indels 73; Gaps 12;

QY 305 TGLHDEVLVVRATQ--LQKNLTCVWGPSTSPKMLSLKLENKAKYSKREKPVWLN 362
DB 91 SGLVLTSSQLTLPAQCLAGKSVTCHVGHYTNP----- 124

QY 363 PEAGMOCCLSDSGQVLESNIKVLPTWSTPVPCCPAPPEKSCDKHTCELLGSPVLF 422
DB 125 -----SQDVTVPQVPSTPTSPSTPTSPSCCH-----PRLSLH 161

QY 423 PPKKDTLMISRTPEVTCVVVDVSHEDPEVKFNNVYDGVENHAKTKPREEOYNSTRYV 482
DB 162 RPLLED-LLLGSAFLITCTLGL-RDASGVTFWTWSSGK--SAGVGPREDICGYSYS 217

QY 483 SVLTVAHODMLNGEKYCKVSNKALPAPIEKTISKAKGPREPOVYTLPPSRDELTKN- 541
DB 218 SVLPSCAEFPMNHCKTCTTAAYPESEKTPPLATLSKS-GMTFREYHLLPPSESELANEL 276

QY 542 VSLTCLVKGFPSPDIIVEMESNGQ--PENNYKTPPVLD-SDG-SFLYSKLTVDKSKW 596
DB 277 VTLTCLARGSPKDVIVRWLQSGQELPREKYLTMASROBPSQGTTFVATSLIRVAEDW 336

QY 597 QQGNVSCSVMEALHNHYTKSLSPG-----LQDENC 632
DB 337 KKQDTPSCWVGHEALPLAFTQKTIIDRLAGKPHVNVSVMAEDGTC 383

RESULT 66
080217 PRELIMINARY; PRT; 487 AA.

AC 080217;
DT 01-JUN-2003 (TREMBLrel. 24, Created)
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypoetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Colon;
RA Strausberg R.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC049143; AA49143.1; -
DR InterPro; IPR003599; IG.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003597; IG-cl.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; IG_3.
DR SMART; SM00409; IG_3.
DR SMART; SM00407; IG1; 3.
DR SMART; SM00406; IG1; 1.
DR PROSITE; PS50835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 2.
KW Hypoetical protein.
SQ SEQUENCE 487 AA; 53019 MW; 31F2C893900A4D80 CRC64;

Query Match 8.3%; Score 282.5; DB 11; Length 487;
Best Local Similarity 22.9%; Pred. No. 3.5e-13;
Matches 131; Conservative 74; Mismatches 197; Indels 169; Gaps 29;

QY 121 LVFGLRANSDTHLLOGSLTLLS-----PPGSSPSVQCRSPRGNIQGGKTLVSQDL 174
DB 1 MYGLNCVFIVFLTKGVSEVKLEAGGGLVQPGSKUSCAA-----SGFTF----- 48

QY 175 ELQDSGTWCTVLQN-QKVEFPIKIDIVLAFQKASSIVYKKEGQVEFSPILAFTEKLT 233
DB 49 -----SNVMMWNVQSPKGLAEWVAEIRLRNNNATYHAEVSKR-----FTISRQD 95

QY 234 GSGELMWQ-----AERASSSKSTFEDLKNKEYSKRYTQDPKLQMGKKLP 279
DB 96 SKGSYVLQMMNIRADBTGIYCTRRGYDDPNMY-FDWAGATTV-TVSEF---AREPT 149

QY 280 LH-LTLPOALP-----QVAGSGLTLALAKTKGLHDEVNLVVRATQLOKNTL 327
DB 150 IYPLTFPQALSDPVIIGCLIHDPYSGTM-----NVT 182

QY 328 CEWGPPTSPKMLSLKENKAKYSKREKPVWLN-PEAGMOCCLSDSGQVLESNIKVL 386
DB 183 ---WG-----KSGKDIITVVFPPA-----LASGRYTMSSQL-T 212

QY 387 LPTWSTPVPCCPAPPEKSCDKTH-----TCPELLGSPVFLPPEPK----- 427
DB 213 LPA---VECEGESVKSQVGHDSNPVQELVNCGICSPFT---PPPSQPSLSIOR 265

QY 428 ---DTLMISRTPEVTCVVVDVSHEDPE-VKFNMYVDGVENHAKTKPREEOYNST---YR 480
DB 266 PALLEDLLGSDASITCTLNGL--RDPEGAVFTW-----EPSTGDAVQKAVQNSCGYS 318

QY 481 VSVLTVAHODMLNGEKYCKVSNKALPAPIEKTISKAKGPREPOVYTLPPSRDELTKN 540
DB 319 VSSVLPSCAEFPMNHSASFEKCTVTHPSDT-LTGTIAKYVTNFPQVHLLPPSESELAN 377

QY 541 Q-VSLTCLVKGFPSPDIIVEMESNGQ--PENNYKTPPVLDSDG--SFLYSKLTVDKSK 594
DB 378 ELVSLTCLVRANPKEVIVRWLHNGNEILSPESYLVFEPKKEGEGATTLVSLRVSAE 437

QY 595 RMQGNVSCSVMEALHNHYTKSLSPG 625
DB 438 IWKQGDQYSCWVGHEALPMNFTQKTIIDRLSG 468

RESULT 67
08WY24 PRELIMINARY; PRT; 497 AA.

AC 08WY24;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE SNCG6 protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Zheng S., Shao X., Cao J., Geng L., Fang Y., Dong Q.;
RT "Identification and characterization of SNCG6, a Ig-like gene which is
RT down-regulated in colorectal cancer.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF283666; AAL36987.1; -
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; IG_4.
DR SMART; SM00406; IG1; 1.
DR SMART; SM00405; IG1; 1.
DR PROSITE; PS50835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 1.
SQ SEQUENCE 497 AA; 53665 MW; F24D08DFA5A63E5 CRC64;

Query Match 8.3%; Score 282.5; DB 4; Length 497;
Best Local Similarity 25.3%; Pred. No. 3.6e-13;
Matches 117; Conservative 60; Mismatches 205; Indels 81; Gaps 20;

QY 223 VKLKGSGELMWQARASSSKSMITFDLKNKEYSVK---RVT--ODPKLQMGKKLP 283
DB 56 VQAPRGQGLEW-----MGMMNPQGTGNTFFAQKFGKRLPSRDTSLNTAYMVLSSIS 106

QY 284 LPQALPOVAGSGL-----TLAEAKTKGLHDEVNLVVRATQLOKN- 325
DB 107 TEDSAIYFCARNTLNGRGFGVNWFPDMCHGTLTVVSASFPNPKVFLSLCSTQPDGNY 166

```

OY 326 -LTCEVMG--PTSPKLMSTLKENKAKVSKREKPVWVLANPEAGMOCCLSDSGVLLS 382
DB 167 VIACLVGFPPQEP---LSVTWSESGGVATARNP-----PSQASGLVYTTSSQTLIPA 218
OY 383 -----NIK-VLPWSTVPCCPAPRPSKCDKHTHCPPELLG--PSVFLPPPKP 426
DB 219 TOCLAGSVTCHVGHVNTNPQDVTVPVCPSTPTPTSPSTPTPTSPSCCHRLSLHRLPAL 278
OY 427 KDTLMISTPTEVTCVVDVSHEDPEVKWVVDVEVHNATKPREQYNSTYRVSVLT 486
DB 279 ED-LHSGEANLTCTLTGL-RDASGVPTFTWTPSSGK--SAVQGPERRDLCCGYSVSVLP 334
OY 487 VLHODWLNKREYKCKVSNKALPAPIEKTISKAKOPREPOVYLTTPSRDELTKNQ-VSLT 545
DB 335 GCAEPMNHGKFTCTTAAPESKYLTPATLPKS-GNTPREPHLLPPSEELALNELVLT 393
OY 546 CLVKGFPSPDIAVEMESNGQ--PENNYKTPPVLID-SDG--SFPLYSKLTVDKSRMOQGN 600
DB 394 CLAGFSPKQVLMVWLGSGQLPREKYLTMASRQEPSCGTTTFAVTSILRVAAEDMKGD 453
OY 601 VFSCSVNHEALHNHTQKSLSLSPG-----LQLETC 632
DB 454 TFSQWVGHEALPLAFTOKTIDRLAGKPTNVNVSVMAEVDGTC 496

RESULT 68
ID Q29028 PRELIMINARY; PRT; 99 AA.
AC Q29028;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE CDA, allele 2 (Fragment).
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.
ON NCBI_TaxId=96823;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=#1183; TISSUE=Blood;
RA MEDLINE=93329116; PubMed=8335933;
RA Gustafsson K., Germana S., Sundt T.M., Sachse D.H., LeGuern C.;
RT "Extensive allelic polymorphism in an exposed region of the
RT mltature. . . . .";
RL J. Immunol. 151:1365-1370(1993).
DR EMBL; X65630; CAA46584.1; -.
DR PIR; I47132; S21462.
DR HSSP; P01730; ICDY.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00406; Ig; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
FT NON_TER 1
FT NON_TER 99
SQ SEQUENCE 99 AA; 11390 MW; C229B5BEA228318F CRC64;

Query Match 8.2%; Score 280.5; DB 6; Length 99;
Best Local Similarity 56.1%; Pred. No. 5.1e-14;
Matches 55; Conservative 19; Mismatches 23; Indels 1; Gaps 1;

OY 32 KGGTVELTCTASOKSIQFHWKSNQIKILGNOSFLTKGP-SKLNDRADSRSLMDQG 90
DB 1 KAGGLAELPCHSSQSKNLPFSWKNSDOIKILSRHNLHMKASVTELSRLDSKKMMWDHG 60
OY 91 NFPLIINKLIKIEDSDTYICEVEDKSEYQLLVFGLTAN 128
DB 61 SFPLIKLEVTDSGIYICEVEDKRIEVLVFLRLTAS 98

RESULT 69
O8VCX4
Q8VCX4

```

```

ID O8VCX4 PRELIMINARY; PRT; 489 AA.
AC O8VCX4;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DE 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN IGH-VJ558 OR A1893585.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
ON NCBI_TaxId=10090;
RX [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon;
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC018322; AAH18322.1; -.
DR MGD; MGI:96486; Igh-VJ558.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003066; IG_MHC.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; Ig; 4.
DR SMART; SM00406; Ig; 1.
DR PROSITE; PS50835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 2.
KM Hypothetical protein.
SQ SEQUENCE 489 AA; 53208 MW; CC85B1194DAFE2C CRC64;

Query Match 8.2%; Score 278.5; DB 11; Length 489;
Best Local Similarity 23.9%; Pred. No. 7.2e-13;
Matches 137; Conservative 71; Mismatches 191; Indels 175; Gaps 32;

OY 118 VOLIVFGLR--NSDTHLD-GQSLTTLSPSSGSPVOCSPKGNIGKGLSVSL 174
DB 6 VFPLFLSVTAGVHSKVLQOSGAEV-----KPGASVKSCKA-----SGYTFSDYFI 53
OY 175 E-LQSGT-----W-----TCTVLQNKVKERKIDIVLAFKASSIVYKKGGEQVEFS 222
DB 54 HMKQSGGLEWIGWNPSSGSIKFNK--FK-DATLTADKSTTV----- 99
OY 223 FPLAFTEKLTGSGELWQAER-----ASSSKWTFIDKKNREVSVK-RVTDPKLQMG 275
DB 100 ----MDLSRLTSDSAVFPARHEDKNGVSLAMFVWQGLVTVSAPAPAEPTI--- 152
OY 276 KKLPLHLTPQALP-----QYAGSGLTLAEAKTGKLDHDEVNLVMPATQLQK 324
DB 153 ----YPLTFPQALSDPVIIGCLIHDFPFGTM----- 181
OY 325 NLTCFVWGPSTPPLMLSLKENKAKVSKREKPVWVLAN-PEAGMOCCLSDSGVLLSN 383
DB 182 NVT--WG-----KSGKDIITVNFPPA-----LASGGRYTMSSQ 212
OY 384 IKVLPWSTVPCCPAPRPSKCDKTH-----TCPELLGSPVFLPPPKP----- 427
DB 213 L-TLPA-----VECPREGESVCSVQHDSNPQVELANNCFGICSPPT---PPPSQPSLS 264
OY 428 -----DTLMIISTPTEVTCVVDVSHEDPE-VKFNMYVDGVEVHNATKPREQYNST- 478
DB 265 LQRPALLEDLLGSDASITLTNGL--RDPEGAFTW-----EPSTGDAVQKAVQNSCG 317
OY 479 -YRVSVGLTVLHODWLNKREYKCKVSNKALPAPIEKTISKAKQPREPOVYLTTPSRDEL 537
DB 318 CYSVSVLTPGCARWMSGSAFKCTVTHPESDT-LTGTIKAVTVNTPPPQVHLLPPSEEL 376
OY 538 TKNQ-VSLTCLVGFSPDIAVEMESNGQ--PENNYKTPPVLIDSG--SFPLYSKLTIV 591
DB 377 ALNELVSLTLVAAPFKVLYVWMLGNELSESLVLEPPLKEPREGATTIVTSVLRV 436
OY 592 DKSRMOQGNVSCSVNHEALHNHTQKSLSLSPG 625
DB 437 SAEIWKQGDQYSCVVGHEALPMNFTOKTIDRLSLG 470

```

```

RESULT 70
091WT1 PRELIMINARY; PRT; 481 AA.
AC 091WT1;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypochemical protein.
OS Mus musculus (mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon;
RA Strauberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC013490; AH13490.1; -.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; IG; 4.
DR SMART; SM00406; IGv; 1.
DR PROSITE; PS50835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 2.
KM Hypochemical protein.
SQ SEQUENCE 481 AA; 52105 MW; 97DF68D159463F65 CRC64;

Query Match 8.1%; Score 276; DB 11; Length 481;
Best Local Similarity 21.9%; Pred. No. 1.le-12;
Matches 140; Conservative 67; Mismatches 207; Indels 226; Gaps 25;

QY 15 QALLLPRAAGNVLGKGDYELTCTAS--OKSIOFPMKNSNOIKLGNQ----- 66
DB 20 QVOLLQSGPE---LVKPGASVKSICKASGYFTSYTHW---VKORPGGLVWIGM 69
QY 67 SFLTKGPSKLNDRADSRSLM---DQGNFLLIKNLKIEDSDTYICEVDQKEVOLVLF 123
DB 70 IYDGDNTKYNKFKKKTITLADKSSSTAYMFLSLTSEBSAVYFC----- 115
QY 124 GLTANSDTHLLQGSLLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSGTWT 183
DB 116 -TRGGGWAFDYMGQGTTLTVSSSEPARPT----- 143
QY 184 CTYLQKQKVEFKIDIVLAFQKASSIVYKKEGEQVEFSPLAFYVEKLTGSGELMWQAE 243
DB 144 -----IYPLTFPQALSSDPYIIGCLLHDYFP-----SGTNNVTW--- 177
QY 244 RASSSKSWITFDLKNKEVSVKRVTOPDKLQMGKKLPLHLTLPOALPOYAGSGNLTL-ALE 302
DB 178 -GKSGK-----DITVNFPPALASGGR-----YTMSSQLTLPAYE 211
QY 303 AKTGKLGHEVNLVVMRATOLQKMLTCEWGPSTPKMLSLKLENKAKVSKREKPVVILN 362
DB 212 CPEG-----SSVCSVQHDSPVQELN 233
QY 363 PEAGMNOCLSDSGQVLESNIKVLPTWSTPVPCP-----APEKSCDKHTTCPELLG 415
DB 234 -----VNCFGICSPPTPPPPSCQ----- 252
QY 416 GPSEVLPFPKPKDITLMISRTPEVTVGVVDVSHEDPE-VKENMYVVDGVEVNAKTKPREBO 474
DB 253 -PGLSLORPALLE-LIGSDASITCTINGL--RDPEBAVFTW-----EPSTGDVAQKKA 303
QY 475 YNST--YRVVSVLTVLHQMNGKEYCKVSNKALPAPIEKTISRAGQPREPOVYTLF 511
DB 304 VQNSCGYSVSVLPGCAERWNSGASPKCTVTHPESDT-LTGTLAKTVAVTPPQVHLFP 362
QY 532 PSDELTKQO-VSLTCLVKGFPYSDIAVWESNQ--PENNYKTTTPVLDSNG--SFL 585
DB 363 PSEBELALNELVSLTCLVRAFNPKEVLYRWLHGNBELSPESYLVFPELKEPGEATYLV 422
QY 586 YSKLTVDKSRWQGNVSCSVMEALHNHTOKSLSPG 625

```

```

DB 423 TSVLRVSAEIMKQGDQYSCWGHFALPNNFTQKTIIDRLSG 462

RESULT 71
08K172 PRELIMINARY; PRT; 482 AA.
AC 08K172;
DT 01-OCT-2002 (TREMBLrel. 22, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DE 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Similar to expressed sequence A1893585.
OS Mus musculus (mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast tumor;
RA Strauberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC028249; AH28249.1; -.
DR PIR; F33932; F33932.
DR PIR; PH1105; PH1105.
DR PIR; PH1108; PH1108.
DR PIR; PH1114; PH1114.
DR PIR; PH1118; PH1118.
DR PIR; PH1119; PH1119.
DR PIR; PH1125; PH1125.
DR PIR; PH1126; PH1126.
DR PIR; PH1128; PH1128.
DR PIR; PH1131; PH1131.
DR PIR; PH1134; PH1134.
DR PIR; PH1139; PH1139.
DR PIR; PH1142; PH1142.
DR PIR; PH1149; PH1149.
DR PIR; PH1150; PH1150.
DR PIR; PH1151; PH1151.
DR PIR; PH1152; PH1152.
DR PIR; PH1153; PH1153.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003597; IG_c1.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; IG; 4.
DR SMART; SM00407; IGc1; 3.
DR SMART; SM00406; IGv; 1.
DR PROSITE; PS50835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 2.
SQ SEQUENCE 482 AA; 52121 MW; A06FF083E771D084 CRC64;

Query Match 8.1%; Score 276; DB 11; Length 482;
Best Local Similarity 22.0%; Pred. No. 1.le-12;
Matches 137; Conservative 68; Mismatches 203; Indels 214; Gaps 24;

QY 30 LGKGDYELTCTAS--OKSIOFPM---KNSNOIKIKLQGSFLLTKGPSKLNDRADSR 84
DB 30 LVKPGASVLSCKASGYFTSYTHWVVKORPGKLEWIGRIDP--NSGGTKYNEKFKSA 87
QY 85 SLM---DQGNFLLIKNLKIEDSDTYICEVDQKEVOLVGLTANSDTHLLQGSLLT 141
DB 88 TLTVDKPSSTAYMQLSLTSEBSAVYYCTRE-----GDYAMQYMGQTSV 133
QY 142 TLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSGTWTCTYLQKQKVEFKIDIV 201
DB 134 TVSSSEPARPT-----IYP 147
QY 202 LAFQKASSIVYKKEGEQVEFSPLAFYVEKLTGSGELMWQARASSSKSWITFDLKNKEV 261
DB 148 LTFPQALSSDPYIIGCLLHDYFP-----SGTNNVTW-----GKSGK----- 183
QY 262 SVKRVTOPDKLQMGKKLPLHLTLPOALPOYAGSGNLTL-ALEAKTGKLGHEVNLVVMRAT 320

```

```

Db 184 DTTNFPALASGR-----YTMSSQLTLPAVECEG-----216
Qy 321 OLQKMLTCEVWGPTSPKMLSLKLENKEAKYSKREKPYVWVLPAGMMQCLLSDSGVLL 380
Db 217 -----ESVKCSYQHNSNPQELN-----234
Qy 381 ESNIKVLEPTWSTPVPCE-----APEBKSCDKHTHCEBLLGSPSVLFPKPKDTLMIS 433
Db 235 -----VNCPCGICSPPTTPPPSCQ-----PSLSIQRLALBD-LILG 269
Qy 434 RTEPVTCTVVDVSHEDPE-VKFNMYVDGVEVNAKTKPREQVNST--YRVSVLTVLH 489
Db 270 SDASITCTLNLG--RDEGAVFTW-----EPSTGKDAVQKAVQNSCGCYSSVSLPGCA 322
Qy 490 QDMNGKEKCKVSNKALPAPIEKTISKAKGPREPOVYTLPPSRDELTKNQ-VSLTCLV 548
Db 323 ERNMSGASFCTVTHPESDT-LTGITAKVYVNTTPPOVHLPPSEELALNELVSLTCLV 381
Qy 549 KGFYPSDIAVWESNGQ---PENNYKTPPVLDSG--SFLYSKLTVDKSRMOQGNVFS 603
Db 382 RAFNPKVLYVWMLHGNBELSPESYLVPFELKEPEGATTYLVTSVLVSAEIMWQSDQYS 441
Qy 604 CSVHSEALHNHYTOKSLSLSPG 625
Db 442 CMVGHEALPMNFTOKTIDRLSG 463

```

RESULT 72

```

Q99KA4 PRELIMINARY; PRT; 487 AA.
ID Q99KA4
AC Q99KA4;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strauberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC004786; AAH04786.1; -
DR HSSP; P01810; 2FBJ
DR InterPro; IPR007110; I9-1like.
DR InterPro; IPR003006; I9_MHC.
DR InterPro; IPR003596; I9_V.
DR Pfam; PF00047; I9; 4.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PSS0835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 2.
KW Hypothetical protein.
SQ SEQUENCE 487 AA; 52554 MW; 7DC8E96DB333077B CRC64;

```

Query Match 8.0%; Score 274.5; DB 11; Length 487;
 Best Local Similarity 21.7%; Pred. No. 1.4e-12;
 Matches 141; Conservative 81; Mismatches 221; Indels 207; Gaps 27;

```

Qy 1 MNRGVPRHLILV---QLALLPATQGNKVLGKSGDTVELTCTAS--QKSIQFTW-- 53
Db 1 MNFPLSLFLVLVLYKGVQCEVQVLESGLV---KPGSLKLSGMAAGFTPTSSVAMWVR 57
Qy 54 ---KSNQIKILNQGSFLTKGSPKLNDRADSRSLMDQGNFPLIKNLIKIEDSDTYIC 109
Db 58 QTPEKRLKEMVATISDGSY-TYYPDNVYGRFTISRDN-AKNILYLQMSHLKSEDTAYYC 115
Qy 110 EVEQKKEVQLLVGCLTANSDTHLLOGOSLTLTLESFPGSSPSV-QCSRGRKRTIOGKT 168
Db 116 ----ADMGSPYGGYSRFD---YWGCGTTIVTSSSARNPITVPLTPPA----- 159
Qy 169 LSVQLBELQDSGTCTVQLNQKKEVEFKIDIVLAFQKASSIVYKKEGEQVEFSFLPAFT 228

```

```

Db 160 -----LSDDPVII-----GCLIHDFP-----176
Qy 229 VEKLTGSGELMWQAEKASSKSWITPDLKNKEVSVKRVQDPKQLQMGKLLPLHLTLPQAL 288
Db 177 -----SGTMNVYWGSKSDITTT-----VNFPAL 200
Qy 289 PQ---YAGSNLTL-ALFAKTKLHDEVNLVWRA-TOLQKMLTCEVWGPTSPKMLSLKL 344
Db 201 ASGGGYMSSQLTLPANEC-----219
Qy 345 ENKEAKYSKREKPYVWVLPAGMMQCLLSDSGVLLSESNIKVLEPTWSTPVPCEAPEKSC 404
Db 220 -----PEGESVKCSYQHDNAVDELVKCS---GPPPCD-PCPPSC 257
Qy 405 DKHTHCEBLLGSPSVLFPKPKDTLMISTPVTCTVVDVSHEDPEVKRWVVDGVEH 464
Db 258 H-----PSLSIQRLALBD-LILGSDASLTCTLNLGRNPEGAV-FTW-----EPS 299
Qy 465 NAKTKPREQVNST--YRVSVLTVLHODMNGKEKCKVSNKALPAPIEKTISKAKQ 521
Db 300 TGDKDAVQKAVQNSCGCYSSVLPGCAERMSGASFCTVTHPESDT-LTGITAKTVN 358
Qy 522 PREPOVYTLPPSRDELTKNQ-VSLTCLVKGFPYSDIAVWESNGQ---PENNYKTPPVL 577
Db 359 TFPPOVHLPPSEELALNELVSLTCLVRAFNPKEVLYVWMLHGNBELSPESYLVPFELKE 418
Qy 578 DSDG--SFLYSKLTVDKSRMOQGNVFSYVHSEALHNHYTOKSLSLSPG 625
Db 419 PGEGATTYLVTSVLVSAETWQSDQYS CMVGHEALPMNFTOKTIDRLSG 468

```

RESULT 73

```

Q8VEAO PRELIMINARY; PRT; 484 AA.
ID Q8VEAO
AC Q8VEAO;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strauberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC019425; AAH19425.1; -
DR MGD; MGI:96486; Igh-VJ558.
DR InterPro; IPR007110; I9-1like.
DR InterPro; IPR003006; I9_MHC.
DR InterPro; IPR003596; I9_V.
DR Pfam; PF00047; I9; 4.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PSS0835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 2.
KW Hypothetical protein.
SQ SEQUENCE 484 AA; 52859 MW; F2522DFF5ED288A6 CRC64;

```

Query Match 8.0%; Score 274; DB 11; Length 484;
 Best Local Similarity 24.8%; Pred. No. 1.6e-12;
 Matches 102; Conservative 58; Mismatches 153; Indels 98; Gaps 20;

```

Qy 289 Poyags--GNLTALBAKTKLHQEVNLVWRA-----TOLQKMLTCEVWG----- 332
Db 79 PEYSASVQGRFTISRNSQNIILYLNW--TLRAEDATYYCARAYSNYYFDNMNGQSTTLT 136
Qy 333 -----PTSPMLSLKLENKAVY-----SKREKPYVWLN-PEAG 366
Db 137 VSSEPAAREPITVLTFPQALSSDPVILIGCLIHDFPSGTWNTVWGSKGDITVNPFA- 195
Qy 367 MMQCLLSDSGVLLSESNIKVLEPTWSTPVPCEAPEKSCKTH-----TCPELLG 416

```

```

Db 196 -----LASGRYTMSQL-TLPA-----VECPBESVKCSVQSDNSNVQELANNCRGICSP 245
QY 417 PSVFLPFPKPK-----DTLMISRTPEVTCVVVDVSHEDPE-VKFNMYVDDGVAV 463
Db 246 PTT---PPPPSCQPSLQRPALIEDLLGSDASITCTLNGL--RDEGAVFTW-----EP 295
QY 464 HNAKTPREEQNST---YRVSVLTVLHODMLNGEKYCKVSNKALPAPIEKTSKAG 520
Db 296 STGKDAVQKAVQNSCGCYSSVSLPGCAERNMSGASFCTVTHPESDT-LGTIAKTV 354
QY 521 QPREPOVYTLPPSRDELTKNQ-VSLTCLVKGFYPSDIAVEWESNGQ---PENNYKTPPV 576
Db 355 NTFPPQVHLPPRSEELALNELVSLCIVRAFNPKEVLVRLMNGNELSPESYIVPEPLK 414
QY 577 LQSDG--SFFLYSKLTVDSKRWQGVFSCVSHHEALHNHYTQKSLSLSPG 625
Db 415 EPEGATYLVTSVLRVSAEIMWQDQYSCMVGHEALPMNFTQKTRLSG 465

```

RESULT 74

```

ID 091WT3 PRELIMINARY; PRT; 481 AA.
AC 091WT3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN IGH-VJ558 OR A1893585.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon;
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL, BC013488; AAI13488.1; -.
DR MGD; MGI:96486; Igh-VJ558.
DR InterPro; IPR007110; IGH-like.
DR InterPro; IPR003006; IGH_MHC.
DR InterPro; IPR003596; IGH_V.
DR Pfam; PF00047; Igh_4.
DR SMART; SM00406; Igh; 1.
DR PROSITE; PS50835; IGH_LIKE; 4.
DR PROSITE; PS00290; IGH_MHC; 2.
KW Hypothetical protein.
SQ SEQUENCE 481 AA; 52022 MW; 4EB5C253038B718 CRC64;

```

```

Query Match 8.0%; Score 272; DB 11; Length 481;
Best Local Similarity 22.6%; Pred. No. 2.2e-12;
Matches 123; Conservative 69; Mismatches 168; Indels 184; Gaps 26;

QY 182 WTCTV-----ONOKVFEKIDIVLAFQKASSIYKKEGVEFSPLA---FT- 228
Db 3 WSCIILFLVATANGVNSQVQLQ-----QPGAEIV--RPGASVKSCTKSTGTFTD 50
QY 229 -----VEKLTGSGELWQA-----ERASSSKSWITFDLKNKEVS 262
Db 51 YMMNWVKQRPQGQGLWEMIGALDPDSYTSYNGKFKGTTLVDTSSSAYMLLSLTSEDSA 110
QY 263 YKAVTDDPKLQMG-----KKLPLHLTLPLALP-----QYAS 294
Db 111 VYFCAGAPRDSYSGYWGQGTTLTVSSBPAREPTIYPLTFPQALSSDPVLIIGCLIHDPFS 170
QY 295 GNLTLLEAKTGHLQEVNLVVMRATQLOKNTLCFCEWGPSPKLMSLKLENKAIVSR 354
Db 171 GTW-----NVT---WG-----KS 180
QY 355 EKEVWYLN-PEAGMOCLLSDSGQVLLSNIKVLPTWSTPVPKPAPEKSCDKTH----- 408
Db 181 GKQITTVNFPDPA-----LASGRYTMSQL-TLPA-----VECPBESVKCSVQSDNSNV 229

```

```

QY 409 -----TCBELIGSPVFLFPPKPK-----DTLMISRTPEVTCVVVDVSHEDPE 451
Db 230 QELNVNCGICSPPTT---PPPPSCQPSLQRPALIEDLLGSDASITCTLNGL--RDE 284
QY 452 -VKFNMYVDDGVAVHNAKTPREEQNST---YRVSVLTVLHODMLNGEKYCKVSNKAL 507
Db 285 GAVFTW-----EPSTGDAVQKAVQNSCGCYSSVSLPGCAERNMSGASFCTVTHPES 339
QY 508 PAPIKTSKAGQPREQVYTLPPSRDELTKNQ-VSLTCLVKGFYPSDIAVEWESNGQ- 565
Db 340 DT-LGTIAKTVTVTFPPQVHLPPRSEELALNELVSLCIVRAFNPKEVLVRLMNGNE 398
QY 566 ---PENNYKTPPVLDSDG--SFFLYSKLTVDSKRWQGVFSCVSHHEALHNHYTQKSLS 621
Db 399 LSPESYIVPEPLKEPBGATTVLVTSVLRVSAEIMWQDQYSCMVGHEALPMNFTQKTR 458
QY 622 LSPG 625
Db 459 RLSG 462

```

RESULT 75

```

ID 091X92 PRELIMINARY; PRT; 482 AA.
AC 091X92;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon;
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL, BC011181; AAI11181.1; -.
DR PIR; F33932; F33932.
DR InterPro; IPR007110; IGH-like.
DR InterPro; IPR003006; IGH_MHC.
DR InterPro; IPR003596; IGH_V.
DR Pfam; PF00047; Igh_4.
DR SMART; SM00406; Igh; 1.
DR PROSITE; PS50835; IGH_LIKE; 4.
DR PROSITE; PS00290; IGH_MHC; 2.
KW Hypothetical protein.
SQ SEQUENCE 482 AA; 51865 MW; 312E01F9C1BC7F3C CRC64;

```

```

Query Match 8.0%; Score 272; DB 11; Length 482;
Best Local Similarity 22.1%; Pred. No. 2.2e-12;
Matches 143; Conservative 69; Mismatches 218; Indels 216; Gaps 26;

QY 10 LLLVQLALLPAAQNGKRVYLGKG-----DTVELCTAS--QKKSIOFMKNSNQK 60
Db 4 LALLCLVTFPSGAL--SOVLKESGPDVLAQSQSIITCTVSGFALTSAISWAPQPEK 62
QY 61 ILNGQSFLLTKGPSKLANDRADRSRL-WDQGNFPLIIRNLKIEDSDT--YICEVEDQKEE 117
Db 63 GLEMLGVMTGQVTNNGALSKRLSISKDQSKSQVFLKQNSLQNTDTRAYC-ARDSNVE 121
QY 118 VQLLVFGLTANSDFHLQGSITLTLBSPGSSPSVQCRSPRGKNIQGGKTLVSQLEIQ 177
Db 122 GAMDYW-----GQGTSTVSSSEPAPEPT----- 144
QY 178 DSGTWTCTVLOQKQKVEFKIDIVLAFQKASSIYKKEGVEFSPLAFTVEKLTGSGE 237
Db 145 -----IYPLTFPQALSSDPVLIIGCLIHDPF-----SGTNK 175
QY 238 LMMQAEKASSSKSWITFDLKNKEVSVKRVTDQPKLQMGKQLPLHLTLPLALPQVAGSGL 297

```

```

Db      176 VTM-----GKSGK-----DITTVNPPALASGR-----YTMSSQL 206
Qy      298 TL-ALBAKTKLHOEVNLVVMRATQLOKNTLCEWGTSPKMLSLKENKAQVSRK 356
Db      207 TLRVNECPREG-----ESVKSQVQHDN 228
Qy      357 PWWNLNDEAGMWOCLLSDSGVLLSNIKVLPTSTPVRP-----ABEPKSCDKHTH 409
Db      229 PVQELN-----VNCRGICSPPTTPPSCQ-----253
Qy      410 CPBLLGSPVFLPPPKRDTLMSRTPEVTCVVVDVSHEDPE-VKFNWYVDGVEVHNAKT 468
Db      254 -----PSLSIQPALPD-LILGSDASICTTLNGL--RDPEGAVFTW-----EPSTGKD 298
Qy      469 KPREBOYNST---YRVSVLTVLHODWLNGEYCKYCNKALPAPIEKTISKAKGQPREP 525
Db      299 AVOKKAVONSCGCGSVSSVLPGCERNNSGASFCTYTHPESDT-LGTIAKVTNTTTPP 357
Qy      526 QVYTLPPSRDELTKNQ-VSLTCLVKGFYPSDIAVEMESNGO---PENNYKTTTPPVLDSG 581
Db      358 QVHLPPSRDELALNELVSLTCLVRAFNPKKVELVRLHGNBELSPESYLVPEPLKPERGG 417
Qy      582 --SFFLYSKLTVDKSRWQGNVPSGVMEALHNHYTQKSLSLSPG 625
Db      418 ATYLVTSVLRVSAEIMWQGDQYSCWVGHEALPMNFTQKIDRLSG 463

RESULT 76
Q91WP5 ID Q91WP5 PRELIMINARY; PRT; 479 AA.
AC Q91WP5;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon;
RA Strauberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC013656; AA013656.1; -
DR InterPro; IPR007110; Ig_Like.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig; 4.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS00835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 2.
KW Hypothetical protein.
SQ SEQUENCE 479 AA; 51603 MM; ECB2D087748584F CRC64;

Query Match 8.0%; Score 271.5; DB 11; Length 479;
Best Local Similarity 24.1%; Pred. No. 2.4e-12;
Matches 136; Conservative 72; Mismatches 188; Indels 169; Gaps 32;

123 FGLTANSDTHLLOQSLLTLT-ES-----PPGSSPVQCRSPKNGIQGGKTLSSVQL 176
Db      3 FGLTLILVLTLLKVGQCEVOLVESGGGLVRFSGSLKYSCAA-----SGLTFSNYM-- 53
Qy      177 QDSGTWTCVTLQN-QKVEFKIDIVLAFOKASSIVYKKEGOVEFSPFLAFTVEKLTGS 235
Db      54 -----SM-----VRQSPKRLW-----VAAINSNGNTYSD-----TWKGRFTISRDN 95
Qy      236 GELMWQBRASSSSSWIT-----FDLAKNEVSVKRYTQDPKIQMGKPLH-LTL 284
Db      96 STLVIQ---MSLSSEDAFYCYRGVGFVWAGTAV-TVSSSP-----AAEPTIYPLTF 147
Qy      285 PQALP-----QVAGSGNLTLLAETKTKLHOEVNLVVMRATQLOKNTLCEWGP 333

```

```

Db      148 PQALSSDPVILIGLLIHDPYPSGTM-----NVT---WG- 176
Qy      334 TSPKMLSLKENKAQVSRKPVVNLN-PEAGMWOCLLSDSGVLLSNIKVLPTWST 392
Db      177 -----KSGKIDITTVNPPA-----LASGRYTMSSQL-TLPA--- 207
Qy      393 PVPCPAPBPCKCDKTH-----TCPELLGSPVFLPPPKR-----DTL 430
Db      208 -VECPREGESVKGSVQGDNSNPVQELNVCPCISPTT-----PPPSQPSLSIQPALDEL 263
Qy      431 MISRTPEVTCVVVDVSHEDPE-VKFNWYVDGVEVHNAKTKPREBOYNST---YRVSVLT 486
Db      264 LIGSDASICTTLNGL--RDPEGAVFTW-----EPSTGKDAVOKKAVONSCGCGSVSSVLP 316
Qy      487 VLHODWLNGEYCKYCNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNQ-VSLT 545
Db      317 GCAERNNSGASFCTYTHPESDT-LGTIAKVTNTTTPPQVHLIPPSSEALNELVSLT 375
Qy      546 CLVKGFYPSDIAVEMESNGO---PENNYKTTTPPVLDSG--SFFLYSKLTVDKSRWQGN 600
Db      376 CLVRAFNPKKVELVRLHGNBELSPESYLVPEPLKPERGGATTVLTVSRVSAEIMWQGD 435
Qy      601 VFGSVMEALHNHYTQKSLSLSPG 625
Db      436 QVSCWVGHEALPMNFTQKIDRLSG 460

RESULT 77
Q90WB5 ID Q90WB5 PRELIMINARY; PRT; 482 AA.
AC Q90WB5;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor.
OS Anas platyrhynchos (Domestic duck).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Anseriformes; Anatidae; Anas.
OX NCBI_TaxID=8839;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RA Chan S.W.S.; Middleton D.L.; Lundqvist M.; Marr G.W.; Higgins D.A.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB378701; AKS59279.1; -
DR InterPro; IPR003599; Ig.
DR InterPro; IPR007110; Ig_Like.
DR Pfam; PF00047; Ig; 4.
DR SMART; SM00409; Ig; 3.
DR PROSITE; PS00835; IG_LIKE; 2.
KW Signal.
FT SIGNAL.
SQ SEQUENCE 482 AA; 54624 MM; BDA0B8EC6172AD0B CRC64;

Query Match 8.0%; Score 271.5; DB 13; Length 482;
Best Local Similarity 25.0%; Pred. No. 2.4e-12;
Matches 111; Conservative 69; Mismatches 169; Indels 95; Gaps 19;

8 RHLLVQLALPLPAQGNVVLGKKGDTVELCTASOKKSIQPHMK-----NSNOIK 60
Db      12 RAVFVLQQLG---THMAHQQIGVSGKEYILNC---KKHDKDVMTWKVEYDAGSSAIIIQ 66
Qy      61 ILNQGSPFLTKGPKSLNDRADRSRLMDQGNFPLIINKLIEBDSPTYICEVQKEEVL 120
Db      67 ILAGK---IFKGRAPMSDSSET-----NONSKLKVSNNRISDAGTYICEGSDRNSISL 118
Qy      121 LVFGLTANSDTHLLOQSLLTLT-LESPPGSSPS-----VQCRSPRGKN 162
Db      119 HAVKLTISNGYFLPGDDLELTVMKSPKSPQRPFTTLTNSHNSRVTPVELQNETFO--- 175
Qy      163 IOGKTLVSVQLBLODSTWTCVTLQNOKKVEFKI--DIVLAFOKAS-SIVYKKEGOV 219
Db      176 ---KVALKVKQLPRTDSGTWICMWHSDSPSINENISPNVAVLGFETKTHLERNVAADVSTV 232

```



```

AC Q91207;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strauberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC010324; AAH0324.1; -
DR InterPro; IPR007110; IG_1ike.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_V.
DR Pfam; PF00047; IG; 4.
DR SMART; SM00406; IG; 1.
DR PROSITE; PSS0835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 2.
KM Hypothetical protein.
SQ SEQUENCE 486 AA; 52682 MW; 4FEF835125DA870B CRC64;

Query Match 7.9%; Score 270.5; DB 11; Length 486;
Best Local Similarity 26.6%; Pred. No. 2.9e-12;
Matches 97; Conservative 54; Mismatches 144; Indels 69; Gaps 14;

QY 282 LTLPOALP-----QYAGSGNLTLLAEATGKLGHEVNLVVRATQLOKNTLCEV 330
DB 153 LTLPRALSDPVIIGCLIHDFPSGTMTWTV-GKSGKDTITVPPPALASGGYTMSSQL 211
QY 331 WGPSPKMLSLKLENKEAVSKREKRVWVNLNPEAGMOCGLSDSGVLESNIKVLP 390
DB 212 TLPA-----VECPGESYKCSVQHSNVAQVDELVKCS-- 243
QY 391 STPPCPAPPEPKSCDKHTHCELLGSPVFLPPPKKDTLMISRTPEVTCVVDVSHEDP 450
DB 244 GPPPPCP-PCPPSCH-----PSLSLQRLALED-LILGSDASITCTLNGLRPEG 290
QY 451 EVKENWYVDGVEVNAKTPREBOYNST--YRVSVLTVLHDMNGKEYCKVSKAL 507
DB 291 AV-FTW-----EPSTGDAVQKAVQNSCGYSVSVLPGCAERWNSGASFKVTTPES 344
QY 508 PAPIEKTISKAKGQPREPOVYTLPPSRDELTKNQ-VSLTCLVKGFYPSDIAVEESNQ- 565
DB 345 DT-LTGITAKITVTPFPQVHLPPPSBELALNELVSLTCLVRAFNPEKVLVRLHGNB 403
QY 566 --PENNNYKTPPVLDSDG--SFPLYSKLTVDKSRMOCQNVFSGVMHEALHNHYTOKSL 621
DB 404 LSPESYLVEFLKPEKGEGATTYLVTVLRVSAETWKGQDYSCMVGHEALPMNFTOKTID 463
QY 622 LSPG 625
DB 464 RLGG 467

```

```

RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC031703; AAH31703.1; -
DR InterPro; IPR007110; IG_1ike.
DR InterPro; IPR003597; IG_CL.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_V.
DR Pfam; PF00047; IG; 4.
DR SMART; SM00407; IGCL; 3.
DR SMART; SM00406; IG; 1.
DR PROSITE; PSS0835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 2.
SQ SEQUENCE 488 AA; 53127 MW; 0E3B156E155733F0 CRC64;

Query Match 7.9%; Score 270; DB 11; Length 488;
Best Local Similarity 25.5%; Pred. No. 3.2e-12;
Matches 99; Conservative 47; Mismatches 126; Indels 116; Gaps 19;

QY 282 LTLPOALP-----QYAGSGNLTLLAEATGKLGHEVNLVVRATQLOKNTLCEV 330
DB 154 LTLPRALSDPVIIGCLIHDFPSGTMTWTV-GKSGKDTITVPPPALASGGYTMSSQL 211
QY 331 WGPSPKMLSLKLENKEAVSKREKRVWVNLNPEAGMOCGLSDSGVLESNIKVLP 389
DB 184 WG-----KSGKDTITVNEPPA-----LASGRYTMSSQL-TLPA 216
QY 390 WSTPPCPAPPEPKSCDKHTHCELLGSPVFLPPPKKDTLMISRTPEVTCVVDVSHEDP 427
DB 217 -----VECPGESYKCSVQHSNVAQVDELVKCS--PPPSQPSLSLQRLA 269
QY 428 DTLMISRTPEVTCVVDVSHEDP-VKFNWYVDGVEVNAKTPREBOYNST--YRVSV 483
DB 270 EDLLGSDASITCTLNGL--RDEGANFTW-----EPSTGDAVQKAVQNSCGYSVSS 322
QY 484 VLVTVLHDMNGKEYCKVSKREKRVWVNLNPEAGMOCGLSDSGVLESNIKVLP 542
DB 323 VLPGCERWNSGASFKVTTPHPSDT-LTGITAKITVTPFPQVHLPPPSBELALNELV 381
QY 543 SLTCLVKGFYPSDIAVEESNQ--PENNNYKTPPVLDSDG--SFPLYSKLTVDKSRM 597
DB 382 SLTCLVRAFNPEKVLVRLHGNBLSPELVLPKPEGATTYLVTVLRVSAETWK 441
QY 598 QGNVFSQVMHEALHNHYTOKSLSPG 625
DB 442 QSDGYSCMVGHEALPMNFTOKIDRLSG 469

RESULT 82
Q9DCD9 PRELIMINARY; PRT; 426 AA.
AC Q9DCD9;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Adult male kidney cDNA, RIKEN full-length enriched library,
DE clone:0610041A01, full insert sequence.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=CS7BL/6J; TISSUE=Kidney;
RC MEDLINE=21085660; PubMed=11217851;
RA Kawai T., Shimagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Aikawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K. I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiya H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pebole G., Quackenbush J.,
RA Schirml L.M., Stauble F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,

```

RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Guentrich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Maehima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Wetz C., Whitaker C., Wilming L.,
 RA Wysshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohatsu S.,
 RA Hayashizaki Y.,
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 DR EMBL/ AK002875; BAB22422.1; -.
 DR PIR/ F33932; F33932.
 DR HSPSP, P01810; 2FBJ.
 DR InterPro: IPR007110; IG-like.
 DR InterPro: IPR003597; IG_c1
 DR InterPro: IPR003006; IG_MHC.
 DR Pfam/ PF00047; 1g; 4.
 DR SMART/ SM00407; IGc1; 2.
 DR PROSITE/ PS50835; IG LIKE; 3.
 DR PROSITE/ PS00290; IG_MHC; 2.
 SO SEQUENCE 426 AA; 45819 MW; 56E1275BA48F6B1 CRC64;

Query Match 7.9%; Score 269.5; DB 11; Length 426;
 Best Local Similarity 26.6%; Pred. No. 2.9e-12;
 Matches 97; Conservative 54; Mismatches 144; Indels 69; Gaps 14;

QY 282 LTLPOLP-----QYASGNLTALAEKTKLHGVNLVVMRATQLOKNTCEV 330
 DB 93 LTLPRALSDPVIIGCLIHDPGSGTNVWV--GKSGDITTVNPPALASGGGYTWSQL 151
 QY 331 WGPSTPKMLSLKLENKAKYKREKPVWLNPEAGMOCILSDSGVLLSNIKYLPW 390
 DB 152 TLPA-----VECPGESVYKSVQHDNNAVQELDVKS--- 183
 QY 391 STVPCCPAPPKSCDKTHTPCRELLGSPVFLPPPKKDTLMISRTBEVTCVVVDVSHEDP 450
 DB 184 GPPPCP-PCPPSCH-----PSLISQRPALSD-LLSGDASLTCTLGLNPNPG 230
 QY 451 EVKFNMYVGVENHANKTKREQYNST---YRVSVLTLYLHDMVINGEKYKKNYKNKL 507
 DB 231 AV-FTW-----EPSTGDAVOKKAVONSCGYSVSVLPGCARMWSGASFCTVTHPS 284
 QY 508 PAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQ-VSLTCLVKGFYPSDIAVEWESNGQ- 565
 DB 285 DT-LTGITAKITVNTTPRQVHLLPPSEBELALNELVSLTCLVAEPKVELYRMLHGNEE 343
 QY 566 --PENNYKTPPVLDSDG--SFPLYSKLTVDKSRWQGVNFGSCVMHEALHNHYTKSKLS 621
 DB 344 LSPESYLVEPRLKEPGGATTVLTGSLRVSAELMKQGDQYSQVGHGALPMNFTQKTID 403
 QY 622 LSPG 625
 DB 404 RLSG 407

RESULT 83

Q91XE1 PRELIMINARY; PRT; 480 AA.
 AC Q91XE1;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein (Fragment).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCB1_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAUSBERG R.;
 RA STRAUSBERG R.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL/ BC010798; AAH10798.1; -.

DR InterPro: IPR007110; IG-like.
 DR InterPro: IPR003006; IG_MHC.
 DR InterPro: IPR003596; IG_V.
 DR Pfam/ PF00047; 1g; 4.
 DR SMART/ SM00406; IGv; 1.
 DR PROSITE/ PS50835; IG LIKE; 4.
 DR PROSITE/ PS00290; IG_MHC; 2.
 KW Hypothetical protein.
 FT NON_TER
 SO SEQUENCE 480 AA; 51936 MW; 20B9234BEF2B41ED CRC64;

Query Match 7.9%; Score 268.5; DB 11; Length 480;
 Best Local Similarity 23.7%; Pred. No. 4.1e-12;
 Matches 107; Conservative 61; Mismatches 169; Indels 115; Gaps 20;

QY 254 PDLKREKSVKGVTPDPKQMGKULPLHTLPOLPYAGSNLTALAEKTKL---H 309
 DB 45 FIFSNYSYMWVQTPDEKLEW-----VATISNSGYATHYPSMKGRFTISR 91
 QY 310 QEVNLVVMRATQLOKNTC-----EVWG-----PTSPKMLSLK 343
 DB 92 NAQNTVLQMTSLNSBDTAVVYCTRGDYWFVWCACTTVVSSPAPRPITYLPFOA 151
 QY 344 LENKEAKV-----SKREKPVWLN-PEAGMOCILSDSGVLLSNIK 385
 DB 152 LSSDPVIIGCLIHDPGSGTNVWVWGKSGKDTTVNFPRA-----LASGRTWSSQL- 204
 QY 386 VLPSTVPCCPAPPKSCDKTH-----TPRELGSPVFLPPPKK- 427
 DB 205 TLPA-----VECPGESVYKSVQHDNPNVQELVNNCPGICSPPT--PPPCQPSLSIQ 257
 QY 428 ----DTLMISRTBEVTCVVVDVSHEDP- VKENMYVGVENHANKTKREQYNST---Y 479
 DB 258 RPALSDLLGSDASLTCTLNGL--RDPGAVFTW-----EPSTGDAVOKKAVONSCGY 310
 QY 480 RVSVLTLYLHDMVINGEKYKKNYKNALPAPIEKTISKAKGQPREPQVYTLPPSRDELTK 539
 DB 311 SVSVLPGCAERKMNWSGASFCTVTHPSDT-LTGITAKITVNTTPRQVHLLPPSEBELAL 369
 QY 540 NO-VSLTCLVKGFYPSDIAVEWESNGQ--PENNYKTPPVLDSDG--SFPLYSKLTVDK 593
 DB 370 NELSLTCLVRAFNKVELYRMLHGNEELSPESYLVEPRLKEPGGATTVLTGSLRVSA 429
 QY 594 SRWQGVNFGSCVMHEALHNHYTKSKLSLSPG 625
 DB 430 ETKQGDQYSCVGHGALPMNFTQKTIDRLSG 461

RESULT 84

Q8VCV5 PRELIMINARY; PRT; 481 AA.
 AC Q8VCV5;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCB1_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Breast tumor;
 RA STRAUSBERG R.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL/ BC018455; AAH18455.1; -.
 DR InterPro: IPR007110; IG-like.
 DR InterPro: IPR003006; IG_MHC.
 DR InterPro: IPR003596; IG_V.
 DR Pfam/ PF00047; 1g; 4.
 DR SMART/ SM00406; IGv; 1.
 DR PROSITE/ PS50835; IG LIKE; 4.
 DR PROSITE/ PS00290; IG_MHC; 2.

KW Hypothetical protein.
 SQ SEQUENCE 481 AA; 52326 MW; 52844CS826807143 CRC64;
 Query March 7.6%; Score 259.5; DB 11; Length 481;
 Best Local Similarity 22.9%; Pred. No. 26-11; Indels 119; Gaps 24;
 Matches 124; Conservative 81; Mismatches 217;
 120 LTVGLRANSPDTHLLQGQSLTLTLESP-----PGSSPVSQVCPSPGRKNI-----QGQ 166
 6 IFFLELGGTGVH-----SEIOLQSGPBLVRPGISVAVSCASGYSITDNIYVVKSHG 61
 Oy 167 KTLIS-VSOLQLQSDGWTCTVLQNKVFEKIDIVLAFQKASSIVYKKEGEVFSFPL 225
 62 KSLKEMIGYIDPYNQGS-----SYNQK--FK-GKATLTVDKSSNTAF----- 99
 Oy 226 AFYVEKLTGSGELMWQAEBRASSKSWTFDLKNEVSVK-RVYDDPKLQMKLPLHLTL 284
 100 -MYLNLTLSBDSAFYVCAREWYG-AMPAFWQGTGLTVSASASAPTI-----YPLTL 150
 Oy 285 PQALP-----QVAGSGNLTALAEAKTGKHOENVLVVRATOLQKLTCEWGP 333
 151 PPAISSDPVLIIGCLIHIFPBGIMVTV-GKSGDITTVFPPALASGGRITMSQULTP 209
 Oy 334 TSPKMLSLKLENKAVSKKEKPVVYLPNPAQMOCLSDSGVLLSNIKVLPTWSTP 393
 210 A-----VECPGESVSKSVQHDSPVQELDVNC-----SGP 240
 Oy 394 VPCAPAPKSCDKHTHTPELGGSVFLPPPKDITLMISRTPEVTCVVVDVSHEDPEVK 453
 241 TPPEPIPIPSQ-----PSLSLORPALED-LLGSDASITCTLNGLRNPEGAV- 287
 Oy 454 FNNVVDGVEVNAKTKPREQVNST---YRVSVTLVHODWLNKGEYKCKVSKALPAP 510
 288 FTW-----EESTGDAVQKAVQNSCCGYSVSLPGCAERKMSGASFCTVTHPE-SGT 341
 Oy 511 IEKTSKAKQPREPOVYTLPPSRDELTKNQ-VSLTCLVKGFPSPDIAYEWESNGQ--P 566
 342 LTGTIAKVTVPPOVHLLPPPEBELALNELSLTCLVRAFNFKEVLRMLHGNBELSP 401
 Oy 567 ENNYKTPPVLDSDG--SFPLYSKLTVDKSRMOQGNVFCGVMEALHNHYTQKSLSP 624
 402 ESYLVFEPFLKEPBGATTVTLVTVRSVSETWKGQDYSQVGHGHALPNNFTOKTIDRLS 461
 Oy 625 G-625
 Db 462 G-462

RESULT 85
 07TMK4 PRELIMINARY; PRT; 479 AA.
 AC 07TMK4;
 DT 01-OCT-2003 (TREMBlrel. 25, Created)
 DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Hypothetical protein.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 ON NCBI_TaxID=10090;
 RX STRAIN=C57BL/6J; TISSUE=Breast tumor;
 RX MEDLINE=22386257; PubMed=12477932;
 RA Strausberg R., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh L.,
 RA Datschenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loguclano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Boeak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whitting M., Madan A., Young A.C., Shcherchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smallus D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Breast tumor;
 RA Strausberg R.;
 RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC055905; AAH55905.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 479 AA; 52209 MW; 532DED9D46D0AED CRC64;
 Query March 7.5%; Score 255; DB 11; Length 479;
 Best Local Similarity 23.8%; Pred. No. 4,4e-11; Indels 122; Gaps 20;
 Matches 110; Conservative 61; Mismatches 169;
 236 GELW--WQARRASSKSWTFDLKNEV-----SVK-RVT---ODPK---LQMGKK 277
 49 GDYMMVMVROSPEKLEWVS-QIRNKRYNYETYSVSKRFTISRDSGCVLYQNNL 107
 Oy 278 LP-----LHLTLPOLP-----QYA 292
 108 RPEMDGIYCTVEGMDYGRGISVTVSSSARNPITYPLTLPPALSSDPVIGCLIDYF 167
 Oy 293 GSGNLTALAEAKTGKHOENVLVVRATOLQKLTCEWMPSTPBMKMLSLKLENKAVS 352
 168 PSGTMVTV-GKSGKIDITTVFPPALASGGRITMSQULTPA----- 208
 Oy 353 KREKPVVLPNPAQMOCLSDSGVLLSNIKVLPTWSTPVPAPAPKSCDKHTCP 412
 209 -----VECPGESVSKSVQHDSPVQELDVNC-----SGPTPPPTTISQ----- 250
 Oy 413 ILGGPSVFLPPPKDITLMISRTPEVTCVVVDVSHEDPEVKFNNVVDGVEVNAKTPRE 472
 251 -----PSSLORPALED-LLGSDASITCTLNGLRNPEGAV-FTW-----EPSTGDAVQK 299
 Oy 473 EQNST---YRVSVTLVHODWLNKGEYKCKVSKNALPPIEKTISKAKQPREPOVYT 529
 300 KAVNSCGCYSVSVSLPGCAERKMSGASFCTVTHPE-SGTLTGTIAKVTVPVTPPVH 358
 Oy 530 LPPSRDELTKNQ-VSLTCLVKGFPSPDIAYEWESNGQ--PENNYKTPPVLDSDG--SF 583
 359 LPPPEBELALNELSLTCLVRAFNFKEVLRMLHGNBELSPESLVFEPFLKEPBGATTV 418
 Oy 584 FLYSKLTVDKSRMOQGNVFCGVMEALHNHYTQKSLSPG 625
 419 LVTVSVLRVSAETWKGQDYSQVGHGHALPNNFTOKTIDRLSG 460

RESULT 86
 08K024 PRELIMINARY; PRT; 480 AA.
 ID 08K024;
 DT 01-OCT-2002 (TREMBlrel. 22, Created)
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Similar to expressed sequence A1893585.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 ON NCBI_TaxID=10090;
 RX STRAIN=C57BL/6J; TISSUE=Breast tumor;
 RA Strausberg R.,

RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC029188; AAH29188.1; -.
DR InterPro: IPR003599; IG.
DR InterPro: IPR007110; IG_1like.
DR InterPro: IPR003597; IG_c1.
DR InterPro: IPR003006; IG_MHC.
DR InterPro: IPR003596; IG_v.
DR Pfam: PF00047; IG_4.
DR SMART: SM00409; IG_3.
DR SMART: SM00407; IGc1; 3.
DR SMART: SM00406; IGv; 1.
DR PROSITE: PSS0835; IG_Like; 4.
DR PROSITE: PSS0290; IG_MHC; 2.
SQ SEQUENCE 480 AA; 51645 MW; 8690A63C669CDBED CRC64;

Query Match 7.5%; Score 254.5; DB 11; Length 480;
Best Local Similarity 25.5%; Pred. No. 4.9e-11;
Matches 93; Conservative 55; Mismatches 147; Indels 69; Gaps 13;

QY 282 LTLPLALP-----QYAGSGNLTALAEKTKLHQEVNLYVMRATOLQKNTCEV 330
DB 147 LTLPLALSDPVIIGCLHDYFPGTMMVTV-GKSGKDIITVNFPPALASGGRTYMSQL 205
QY 331 WGPSPKMLSLKLENKAKVSKREKPVVNLNPEAGMOCILSDSGQVLESNIKVLPTW 390
DB 206 TLPA-----VECEGESVKCSVQHDSNPQELDVC----- 236
QY 391 STVPCEPAPEKSCDKTHTCPPELLGSPVFLFPKPKDTLMSRPEVTCVVVDVSHEDP 450
DB 237 SGTPPPPTITPSCQ-----PSLSIQRPALPD-LILGSDASITCTINGLRNPG 284
QY 451 EYKFNWYVDGVVHNNAKTKRREQYNST---YRVSVLTVLHODMLNGEKYCKVSNKAL 507
DB 285 AV-FTW-----EPSTGKDAVQKKAQVNSCGCYSVSVLPCEAERNMSGASFCTVTHPE 337
QY 508 PAPIEKTISKAKQPREPOVYTLPPSRDELTKNQ-VSLTCLVKGFPYSDIAVEMESNQ- 565
DB 338 SGLTGTIAKYVNTFPPQVHLLPPSEBELALNELSLTCLVRAFPKPKVLRMLHGNE 397
QY 566 --PENNYKTPPVLSDG--SFPLYSKLTVDKSRMOQGVFSCSVNHEALHNHYTKSLIS 621
DB 398 LSPESYLVEEPLEKEPEGATTVLVTSVLRVSAETWKQDQYSCMVGHEALPMNFTOKTID 457
QY 622 LSRG 625
DB 458 RLUG 461

RESULT 87
ID Q99M22 PRELIMINARY; PRT; 479 AA.
AC Q99M22;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strauberg R.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC002091; AAH02091.1; -.
DR HSSP: P01810; 2FBJ.
DR InterPro: IPR007110; IG_1like.
DR InterPro: IPR003006; IG_MHC.
DR InterPro: IPR003596; IG_v.
DR Pfam: PF00047; IG_4.
DR SMART: SM00406; IGv; 1.
DR PROSITE: PSS0835; IG_Like; 4.
DR PROSITE: PSS0290; IG_MHC; 2.

KM Hypothetical protein.
SQ SEQUENCE 479 AA; 51992 MW; 768B39A138918692 CRC64;

Query Match 7.4%; Score 253.5; DB 11; Length 479;
Best Local Similarity 25.5%; Pred. No. 5.8e-11;
Matches 93; Conservative 55; Mismatches 147; Indels 69; Gaps 13;

QY 282 LTLPLALP-----QYAGSGNLTALAEKTKLHQEVNLYVMRATOLQKNTCEV 330
DB 146 LTLPLALSDPVIIGCLHDYFPGTMMVTV-GKSGKDIITVNFPPALASGGRTYMSQL 204
QY 331 WGPSPKMLSLKLENKAKVSKREKPVVNLNPEAGMOCILSDSGQVLESNIKVLPTW 390
DB 205 TLPA-----VECEGESVKCSVQHDSNPQELDVC----- 235
QY 391 STVPCEPAPEKSCDKTHTCPPELLGSPVFLFPKPKDTLMSRPEVTCVVVDVSHEDP 450
DB 236 SGTPPPPTITPSCQ-----PSLSIQRPALPD-LILGSDASITCTINGLRNPG 283
QY 451 EYKFNWYVDGVVHNNAKTKRREQYNST---YRVSVLTVLHODMLNGEKYCKVSNKAL 507
DB 284 AV-FTW-----EPSTGKDAVQKKAQVNSCGCYSVSVLPCEAERNMSGASFCTVTHPE 336
QY 508 PAPIEKTISKAKQPREPOVYTLPPSRDELTKNQ-VSLTCLVKGFPYSDIAVEMESNQ- 565
DB 337 SGLTGTIAKYVNTFPPQVHLLPPSEBELALNELSLTCLVRAFPKPKVLRMLHGNE 396
QY 566 --PENNYKTPPVLSDG--SFPLYSKLTVDKSRMOQGVFSCSVNHEALHNHYTKSLIS 621
DB 397 LSPESYLVEEPLEKEPEGATTVLVTSVLRVSAETWKQDQYSCMVGHEALPMNFTOKTID 456
QY 622 LSRG 625
DB 457 RLUG 460

RESULT 88
ID Q9W6V7 PRELIMINARY; PRT; 487 AA.
AC Q9W6V7;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE CD4 protein precursor.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=H.B2; TISSUE=Thymus;
RX MEDLINE=99218434; PubMed=10201936;
RA Koskinen R., Lammiakki U., Tregaskes C.A., Salomonsen J., Young J.R.,
Vainio O.;
RT "Cloning and modeling of the first nonmammalian CD4.";
RT J. Immunol. 162:4115-4121 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=B12;
RA Koskinen R.H., Salomonsen J., Tregaskes C.A., Young J.R.,
"The chicken CD4 gene has remained conserved in evolution.";
RT Immunogenetics 0:0-0 (2002).
DR EMBL: Y12012; CAA72740.1; -.
DR EMBL: AJ401223; CAC82027.1; -.
DR InterPro: IPR007110; IG_1like.
DR Pfam: PF00047; IG_3.
DR PROSITE: PSS0835; IG_Like; 2.
DR KX Signal.
FT SIGNAL 1 28 POTENTIAL.
SQ SEQUENCE 487 AA; 54985 MW; B9CBA92EC9F7F45B CRC64;

Query Match 7.0%; Score 240; DB 13; Length 487;
 Best Local Similarity 25.7%; Pred. No. 6,56-10;
 Matches 113; Conservative 61; Mismatches 164; Indels 82; Gaps 20;

```

QY 10 LLLVQLALLPAATQGNKLVGKKQDVELTCTA-SOKSKIQFMKN-----SNOIKILG 63
D 14 VILVQLQGLTFIMQOEQOI-GIAGKEVILSSKAIINQDGTWKYKKEVSSTIISFS 72
QY 64 NQGSFLTKGPKSKLNDRAISRSLMDQGNFPIIKIKIIESDPTICEVEDQKEVQLLVF 123
D 73 KAQVFKEGAPM-----THRSLSNSKSKLKVSDLSLDAGIYTCACYPVVSISLHV 125
QY 124 GLTANSDTHLQGSILTLT-ESPPGSSPS-----VQCSPPGKNIQ 165
D 126 KLTISSNGHFLTNDELITLQNSHSHOPHSLIKLFNINDIYTELLOEAPQ-KTI-- 182
QY 166 GKTLSVQSLQDSGTWTCTVLQNKVFEKI-DIVLAFQKAS-SIVYKKEGQVEFS 222
D 183 ---LKLQQLKAIQSGTWMCHVSNPSINQISFDVKVIGPEKRELEIYTVGNTAILLS 239
QY 223 PPLAFYKELTG--SGELMWAESAASSKSWITDILKKEVSVKRYQDPKLQNGKLLPL 280
D 240 WRLNFRKI KWKEGFTGKLMEPQGNTHIHELINS-----VTHQELHRTKK-SN 288
QY 281 HL-----TLPOLP--QYAGSGNTLTAELAKTGKLGHEVNLVVMRATOL----- 322
D 289 HIWEISEGKTGDMVKIPVQLNHSQYKQLEINGRTRRESRALVMQVTAIPMGPL 348
QY 323 ---QKNLTCFVWGPTSPKMLSLKLEN---KEAKVSKREKPVVNLPEA-GMOC-LLS 373
D 349 SRGKMTLLCGVSPPLPSNAHLMEVRVNGTQMEKKSKQHAKVAVVNSAPGLNCHLVE 408
QY 374 DSG-QVLLSEIKIKLPTWST 392
D 409 DNNKISLNTVEEAHVNS 428

```

RESULT 89

```

ID P70443 PRELIMINARY; PRT; 90 AA.
AC P70443;
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Mutant T-cell surface glycoprotein CD4 (Fragment).
GN CD4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BR/cdJ;
RA Rosenwasser O.A., Fairchild S.P., Tomonari K.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U75219; AAB18133.1; -.
DR HSSP; P05540; ICID.
DR InterPro; IPR007110; IG-like.
FT NON TER
SQ SEQUENCE 90 AA; 9938 MW; 6528B707928AA4685 CRC64;

```

Query Match 6.9%; Score 237; DB 11; Length 90;
 Best Local Similarity 59.2%; Pred. No. 9,7e-11;
 Matches 45; Conservative 14; Mismatches 17; Indels 0; Gaps 0;

```

QY 310 QEVNLVVRATQLOKNTLCEVWGPTSPKMLSLKLENKEAKVSKREKPVVNLPEAGMWO 369
D 1 QEVNLVVMKVAQLNNTLCEVWGPTSPKMLSLKLENKEAKVSKREKPVVNLPEAGMWO 60
QY 370 CLUSDGQVLLSENIK 385
D 61 CLLSEGDVKVMDRSIQ 76

```

RESULT 90

```

ID Q96RW7 PRELIMINARY; PRT; 5636 AA.
AC Q96RW7;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hemocentin.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Trent J.;
RT "Human hemocentin gene.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF156100; AAK68690.1; -.
DR GO; GO:0005727; C:extrachromosomal circular DNA; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR00152; Asx hydroxyl_5.
DR InterPro; IPR000875; Cectropin.
DR InterPro; IPR001434; DUF11.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF_Ca.
DR InterPro; IPR009017; GFP-like.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IG_c2.
DR InterPro; IPR00169; SHPoc_acsite.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF00047; EGF_5.
DR Pfam; PF00090; tBP_1; 6.
DR SMART; SM00179; EGF_CA_7.
DR SMART; SM00408; IGc2; 43.
DR SMART; SM00209; TSP1; 6.
DR SMART; SM00327; VWA; 1.
DR TIGRFAMs; TIGR01451; B and repeat; 9.
DR PROSITE; PS00010; ASX_HYDROXYL; 5.
DR PROSITE; PS00268; CECTROPIN; 1.
DR PROSITE; PS01187; EGF_CA; 8.
DR PROSITE; PS01187; EGF_CA; 3.
DR PROSITE; PS00835; IG_LIKE; 44.
DR PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
DR PROSITE; PSS0092; TSP1; 6.
KW EGF-like domain; Immunoglobulin domain.
SQ SEQUENCE 5636 AA; 613660 MW; F000B19CED7B52C CRC64;

```

Query Match 5.8%; Score 198; DB 4; Length 5636;
 Best Local Similarity 19.9%; Pred. No. 3,7e-05;
 Matches 157; Conservative 105; Mismatches 252; Indels 274; Gaps 39;

```

QY 34 GDTVELTCTASQKSKIQFHKNSNQIKLGNQGSFLTKGPKLNDRAISRSLMDQGNFP 93
D 807 GSNVTLPCVQYQVPEPTIKMRRLDNNPISR--PFSVSSISQURTGA----- 851
QY 94 LITKLIKIESDPTICEVEDQ---KEVQLLVFGTLA-----NSDTHLQGSILTLT 143
D 852 LFLTNLWASDKGYICEABNQFSKIQSETTVTYTGVLAPLIGISPVANVIEGQOLTLPC 911
QY 144 ESPPGS-----SPSVQCSPPGKNIQGGKTLVSQLELDQSGTWC--- 184
D 912 TLLAGNPIDRRKIKSAMLQNPITYVS-----DGLHTEVQLDGGSEYICVAS 963
QY 185 -----TVLQNKVFEKIDIVLAFOKAS-----SIYKKEGQVEF 221
D 964 NVAGTNKKTTSVVVHVLPTIQHGQILSTEGIPVTLPCASGNPKPSVIWMSKGLIST 1023

```



```

Db 58 QAPGKLEWVSGISWNSGSGICGVADSVKGRFTIS-RDNANKSLYLQNN-----SLRAEDT 110
QY 105 DTVICEDVDQKEEVQLVFLGLTANSDPHLQGSGLTLTLSPSSPSV-----QCRSPR 159
Db 111 ALTYCAHKGSGSYI-----GYTGMND--WMGGTTVTSAPKADVPPIISGCHPK 162
QY 160 GKN-----IQQGKTLVSQLELDQSGTWTCTVLQKQKVEKIDIVLAFKASSIV 212
Db 163 DNSPVLACILTGHPSTV-----TW-----Y 186
QY 213 KKEGEQVFFSFP-----LAFVYKLTGSGELMWQAE-----PSSSKSHITFDLKNKE 260
Db 187 MGTQSPQRTFFPEIQRDSVYMTSSQLSTPLQQRQGEYKCVQVHTASKS-----KKE 239
QY 261 VSVKRVTPDPRLOMGKPLHLTLPLPOLPVAGS---GNLLALEATKGLHQEVNLVW 317
Db 240 IF--RWESPFAQAS-----SVPTQOPQAGSLAKATTAATTATNTGKGE----- 284
QY 318 RATQLQKRLTCEVWGPTSPKMLSLKLENKAKYKREKPVVNLNPEAGMQLSDSGQ 377
Db 285 -----KKEKEKEBOEBET----- 299
QY 378 VLESNIKVLFTMTSPVPCPAPEPKSCDKHTCELLGSPVFLPPPKDMLMISTPE 437
Db 300 -----KTP--ECP-----SHTQF--LG---VYLLTPAVQD-LMLRDKAT 330
QY 438 VTCVVVDVSHEDPEVKNNWYVDG--VEVNAKTRPREQVNSTYVSVLTLYLHODWLNGK 496
Db 331 FTCVVSQSDLKDAHL--TWVAGVAPPGVGEGLLEHNSGSGSHRLTLPRSLMAGT 388
QY 497 EYKCKVANKALP-----APIEKTIS-KAKQPREPVYTLPPSRDELTKQV 542
Db 389 SVCTLNHPSLPPEQLMAKREPAQAQVYKSLNLAASSDPPEASW----- 434
QY 543 SLTCLVKGFPYPSDIAVWESNGQPENN--YKTPPVLDGSGSFYLSKLTVDKSRWQG 599
Db 435 -LICEVGFSPSPNILLMWLEDPQREVTSGFAPARPPQSGTTFWMSVLRVPAPSPQ 493
QY 600 NVFSCSVNHE--ALNHHYTKSLSLSPGLQDFTCAADQGLD-----GLMTT 646
Db 494 ATYTCVSHEDSRTLLN--ASRSLEVS-YLAMPPLIQSKDENSDDYTFDVGSLMTT 549

RESULT 93
Q9D8M4 PRELIMINARY; PRT; 130 AA.
AC Q9D8M4;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE 1810027001Rik protein.
GN 1810027001Rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Pancreas;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Araiawa T., Hara A., Fukunishi Y., Komoto H., Adachi J., Fukuda S.,
RA Aizawa K., Itawa Y., Nishi K., Kiyosawa H., Kondo S., Yamana K.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kodota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fletschmann W., Gaasterland T., Gissi C., King B., Kocha H.,
RA Kuehl L.M., Lewis S., Matsuo Y., Nakai T., Pease G., Quackenbush J.,
RA Schirml L.M., Straub F., Suzuki R., Tomita M., Wagner B., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barh G.,
RA Blake J., Botelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gierlich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaeets P.,

```

```

RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seta Y., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wyszewski-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohseki S.,
RA Hayashizaki Y.,
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK07622; BAB25142.1; -.
DR PIR; S22760; S22760.
DR HSP; P01842; TPAB.
DR MGD; MGI:1916426; 1810027001Rik.
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003597; IG_c1.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; IG_1.
DR SMART; SM00407; IGc1.1.
DR PROSITE; PSS0835; IG_LIKE.1.
DR PROSITE; PSS00290; IG_MHC.1.
SQ SEQUENCE 130 AA; 14253 MW; 438197975E76B54 CRC64;

Query Match 5.4%; Score 183.5; DB 11; Length 130;
Best Local Similarity 36.5%; Pred. No. 2.1e-06;
Matches 38; Conservative 18; Mismatches 43; Indels 5; Gaps 3;

QY 520 GQPR-EPQVYTLPPSRDELTKQVSLTCLVKGFPYPSDIAVWESNGQPENNYKTPPVL 577
Db 25 GQKSSPSVTLFPSSSELETNKATLVCTITDTPGVVYVDMKVDGTPVQGHETTOPSK 84

QY 578 DSDGSFELYSKLTVDKSRWQGVNFCVWHEALNHHYTKSL 621
Db 85 QSNKKWASVYTLTLRAWEHRSYSGQVTHE--GHVYKSL 125

RESULT 94
Q8N355 PRELIMINARY; PRT; 234 AA.
AC Q8N355;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC028090; AAH28090.1; -.
DR PIR; S12441; S12441.
DR InterPro; IPR003599; IG.
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003597; IG_c1.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; IG_2.
DR SMART; SM00409; IG_2.
DR SMART; SM00407; IGc1.1.
DR SMART; SM00406; IG_1.
DR PROSITE; PSS0835; IG_LIKE.2.
DR PROSITE; PSS00290; IG_MHC.1.
KM Hypothetical protein.
SQ SEQUENCE 234 AA; 24792 MW; CC848CABA4A9D63 CRC64;

Query Match 5.4%; Score 183.5; DB 4; Length 234;
Best Local Similarity 25.0%; Pred. No. 4.9e-06;
Matches 55; Conservative 39; Mismatches 99; Indels 27; Gaps 7;

QY 404 CDKTHPCPELLGSPVFLPPPKDMLMISRTPEVTCVVVDVSHEDPEVKNNWYVDG 457
Db 14 CTGSVTSVYLTQPPSVAP-----GQTAIRITCGGNISK-----SVHHYQKQPKQ 60

```

```
QY 458 VDCVEVHNAKTRKPR---EEQYNSTYRVSVLTVLHODMLNGKRYCKKSNKALPAP1---- 511
DB 61 APVLVYVYDSDRSPGIPERFSGSNGSNTATLTISRVDAGGEALYQQLWSSSDHPVFG 120
QY 512 EKTISAKAGQPR-EPQVYTLPPSRDELTKNOVSLTGVKGFYPSDIAYEWESNGQP-ENN 569
DB 121 GGTGKLTVLQGPKAAPSVTLFPSSSELOANKATLVCLISDFYGGATVAMKADSPYKAG 180
QY 570 YKTPPVLDSDGSFFLYSKLTVDKSRMVGQGNVFSQVMHE 609
DB 181 VETTPSKQSNMKYAASSVYLTLPQWKSRSYSQVTHE 220

RESULT 95
Q99M11 PRELIMINARY; PRT; 235 AA.
ID 099M11;
AC 099M11;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN 11)
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC002129; AAH02129.1; -.
DR HSP: P01703; 7FAB.
DR InterPro: IPR007110; Ig_1like.
DR InterPro: IPR003596; Ig_MHC.
DR Pfam: PF00047; Ig_2.
DR SMART: SM00406; IgV_1.
DR PROSITE: PS00835; IG_LIKE_2.
DR PROSITE: PS00230; IG_MHC_1.
KW Hypothetical protein.
SQ SEQUENCE 235 AA; 25403 MW; 39807BFE6782A3FB CRC64;

Query Match 5.4%; Score 183.5; DB 11; Length 235;
Best Local Similarity 24.9%; Pred. No. 5e-06;
Matches 38; Conservative 39; Mismatches 101; Indels 35; Gaps 9;

QY 407 THTCPBLGGPSVFLPPPKKDTLMTSRPEVTCVVDVSHEDPEYKFNMY----- 457
DB 15 TGSCAQQLV-----LHQPSVST-SLGSYAKLPCKASTGNIGDSYV--NMYYQYMGSRPT 65
QY 458 --VDGVEVHNAKTRKPR---EEQYNSTYRVSVLTVLHODMLNGKRYCKKSNKALPAP 510
DB 66 NMYYGDDLRLPSGVSDRSGSIDSSNSAF-----LTQNVQADEDDYCYOSSSGIRVF 120
QY 511 EKTISAKAGQPR-EPQVYTLPPSRDELTKNOVSLTGVKGFYPSDIAYEWESNGQP-EN 568
DB 121 GGTGKLTVLQGPKAAPSVTLFPSSSELETKATLVCTISDFYGGATVAMKADSPYKAG 180
QY 569 NYKTPPVLDSDGSFFLYSKLTVDKSRMVGQGNVFSQVMHEALHNHYTKSL 621
DB 181 GVETTPSKQSNMKYAASSVYLTLPQWKSRSYSQVTHE---GHTVEKSL 230

RESULT 96
Q96AA2 PRELIMINARY; PRT; 6620 AA.
ID 096AA2;
AC 096AA2;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Obscurin.
GN OBSCN.
OS Homo sapiens (Human).
```

```
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN 11)
RP SEQUENCE FROM N.A.
RA MEDLINE=21342081; PubMed=11448995;
RC TISSUE=Heart;
RX Young P.W., Ehler E., Gautel M.;
RT "Obscurin, a giant sarcomeric Rho guanine nucleotide exchange factor
RT protein involved in sarcomere assembly."
RT J. Cell Biol. 154:123-136(2001).
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
DR EMBL: AJ002535; CAC44768.1; -.
DR GenBank: HGNC:15719; OBSCN.
DR GO: GO:0005524; F-ATP binding; IEA.
DR GO: GO:0004812; F-ATPase activity; IEA.
DR GO: GO:0006418; P-amino acid activation; IEA.
DR InterPro: IPR003961; FN III.
DR InterPro: IPR008957; FN_III-like.
DR InterPro: IPR007110; Ig_1like.
DR InterPro: IPR003598; Ig_c2.
DR InterPro: IPR000048; IQ_region.
DR InterPro: IPR001849; PH.
DR InterPro: IPR000219; RhGEF.
DR InterPro: IPR001452; SH3.
DR InterPro: IPR001412; tRNA-synt_I.
DR Pfam: PF00041; tns; 2.
DR Pfam: PF00047; Ig; 49.
DR Pfam: PF00612; IQ_1.
DR Pfam: PF00169; PH; 1.
DR Pfam: PF00621; RhGEF; 1.
DR SMART: SM00060; FN3; 2.
DR SMART: SM00408; IGC2; 17.
DR SMART: SM00015; IQ_1.
DR SMART: SM00233; PH; 1.
DR SMART: SM00326; SH3; 1.
DR PROSITE: PS00178; AA_TRNA_LIGASE_I; 1.
DR PROSITE: PS50010; DH_2; 1.
DR PROSITE: PS50835; IG_LIKE; 46.
DR PROSITE: PS50096; IQ_1.
DR PROSITE: PS50003; PH_DOMAIN; 1.
KW Immunoglobulin domain.
SQ SEQUENCE 6620 AA; 721665 MW; C2AE8EB7B284452 CRC64;

Query Match 5.3%; Score 182.5; DB 4; Length 6620;
Best Local Similarity 19.7%; Pred. No. 0.00072;
Matches 135; Conservative 93; Mismatches 269; Indels 187; Gaps 28;

QY 29 VLGKGGDTVELTCTASOKKSIQFMKNSNQIKLGNQGSFLTKGPKLNDRADSRSLMD 88
DB 1818 VCGELGGVTLTLCGLS-PACAEVVRGCGTQPRVGKRFQMAEGPVR----- 1863
QY 89 QGNFLLIKNLKIEDSDTYICEVEDQKEVQLVFGLTANSDDLHLLQGSLTLTLESPPG 148
DB 1864 ----SLTVLGLAADAAGEVYCESRDDHTSAQLTV----- 1893
QY 149 SPSVQCSPPKAKNIQGGKTLVSQLELSDSTWTCTYLQNKQKVEFKIDIVLAF---- 204
DB 1894 ----SV-----PRVYKMSG--LSTVAEEGEGATPCVSPS-----DVAVWFRDGA 1936
QY 205 --QKASIVYKKEGQVERSPFLAFTVEKLTGSGELMWAQRASSKSWITFDLKNKEVS 262
DB 1937 LLOPSEKPAISQSGA----SHSLTISDLVLEAGQITVEAEQASSAA-----LKVREAP 1987
QY 263 V--KRVTDQPKLQMGKQLPLHLTLPLQALPQVAGSGLTLALBAKTKGHQE---VNLVVM 317
DB 1988 VLFKKLEPQVVEERSSVTLVELTRPWPBELKWTNRATATLAKGVIEIHAEGARHLVLH 2047
QY 318 KATLOLQKN-LTCEWGPSPSKMLSLKLENKAKYSKREKPPVNLNPEKGMQCLSDSG 376
DB 2048 NVGFADRGFGCET---PDDKTQAKLTVMRQVRLVR-----GLQAVEAREQGS 2092
QY 377 QVLE---SNIKVLPWTSTPVPVCPAPBPKSCDKHTCPFLGSP--SVFLPPKPKDT-L 430
```

```

Db      2093 TATHEVOLSHADVGSWTR-----DGLRFQOGPTCHLAVRGPMHTLTLGLRPFDSGL 2145
Qy      431 MISR-----TPEVTCVVDVSHDEBYKFNWVDDEV 463
Db      2146 MVEFAEGVHTSARLVLTLPVFSRPLQDVVTTKEKVTLECELSRNVVDVRMLKDVEL 2205
Qy      464 HNAKTREREQYNSTYRVSVLTVLHODWLNKGEYKCKVSNKALPAPL----- 511
Db      2206 RAGKTMALMAQ-----GACSLTITRCEFPADQGVYVDADHDAQSSASVYKQGRYTLITR 2260
Qy      512 -----EKTISKAGQPREPQVYTLPPSSDELTL--KNQVSLTCLYKGFYPSD 555
Db      2261 RVLAEDAGEIQFVLENAESRAQLRVKELPTLVRLPDKIAMEKRGVLEQVVS---RAS 2317
Qy      556 IAVWESNG---QENNYKTPPTPLDSDGFLLYKSLTVDKSRKQGNVSCSV--MHEA 610
Db      2318 AQVWFKGSEQLQGPVTE---LVSDG---LYRKLLISDVHAEDEDTYTCDAGDVKTS 2369
Qy      611 LHNHYTOKSLSLSPGLQDETCAE 634
Db      2370 AQFVEEQSITIVRGLQ-DVTWME 2392

RESULT 97
Q8ISF3 PRELIMINARY; PRT; 2693 AA.
AC Q8ISF3;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, last annotation update)
DE 301kDa_2 protein.
GN ISOF.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Pelodertinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22269627; PubMed=12381307;
RA Flaherty D., Gernert K., Shmeleva N., Tang X., Mercer K.,
RA Borodovsky M., Benian G.;
RT "Titins in Caenorhabditis elegans with Unusual Features: Coiled-coil
RT Domains, Novel Regulation of Kinase Activity and Two New Possible
RT Elastic Regions."
RL J. Mol. Biol. 323:533-549(2002).
DR EMBL; AY130758; AAN61521.1; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
DR GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.
DR GO; GO:000468; F:protein amino acid phosphorylation; IEA.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IG_c2.
DR InterPro; IPR00719; Prot_kinase.
DR InterPro; IPR002290; Ser_Thr_kinase.
DR InterPro; IPR008271; Ser_Thr_kinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00047; fn3; 2.
DR Pfam; PF00047; fg; 19.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Proc_kinase; 1.
DR SMART; SM00060; FN3; 1.
DR SMART; SM00408; IGc2; 19.
DR SMART; SM00220; S_TKC; 1.
DR SMART; SM00219; TYRKC; 1.
DR PROSITE; PS00835; IG_LIKE; 18.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PSS0011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PSS0108; PROTEIN_KINASE_ST; 1.
SQ SEQUENCE 2693 AA; 299094 MW; 1CE4AD2011E8391A CRC64;

```

Query Match 5.3%; Score 179.5; DB 5; Length 2693;

```

Best Local Similarity 19.8%; Pred. No. 0.00034;
Matches 133; Conservative 106; Mismatches 258; Indels 175; Gaps 32;

Qy      28 VLGKGDVLTCTLSQKKSIFHWKNSQIKLNGQSFLLTKGSKLNDRAISRSLW 87
Db      1100 VVLTAGEITATFTQSNAPAAQVW-----LHNGKALQOTKSNYKTRLF 1144
Qy      88 DQGNFPLILNGLIEDSDPTICEVEDQKEEV---OLLVGLTRANS----- 129
Db      1145 DDNTATILVIBNVDELQGYTAVANNQFGDVHTSAQLTISGSAKTIASLPFIIEIKP 1204
Qy      130 DTHLQGSQSLTLIESPPGSSP-----SVQCRSPR-----GKNIQGGKTLVSQ 173
Db      1205 KINWVG--ATLSIQADLNGSPILPVWLKDNSELVESDRILQKCGVWVQ---LLVRD 1258
Qy      174 LEIQDSGTWTCTYLQNKQKVEFKIDVLAFOQASSTIVYKGEQVVF-----S 222
Db      1259 VGLEDEGTITITANEKGIKQNTVEVS---TSKSVKEKEKKEKDEGKKKGRPG 1315
Qy      223 FPL---AFTVEKLT-----GSGELMMQAEBAASSKSWTF-DLKNKESVYK----- 264
Db      1316 LPRPSGASKTEQVTMAPDAPBSGPAUSVEVERCPQREWVSCGSTSLLEIKGLTPNT 1375
Qy      265 ----RVTOPKLOMGKPLPLHLTPQA---LPQVAGSNLTALBAKTGKLHQEVNLVY 316
Db      1376 EYIFRVAGKXKQGLGEMSEMTSLTKTASVQAQPF-----TISPQK-----IIA 1420
Qy      317 MRATLOKNTLCVWGPTSPKMLSLKENKAVSKREK-----PVWVLN---PEA 365
Db      1421 NRDEFE--IAVEFGSTPTP---SVKMYKENIQIVDEKIDVATTSTSSILMLKQSEN 1474
Qy      366 GMMQCLL-SDSGVULESNIKVLPWTSTPPCAPAPKSDCKHTCEBLLGGSVLPFP 424
Db      1475 GTFNCLLENLGGASASCOVTT---FNKPSLQSTPDHSLERN-----LVPT 1518
Qy      425 KPR----DTLMSRTEVTCVVDVSHDEPEYKFNWVDGEVHNAKTREREQYNSTYR 480
Db      1519 LQKALNNESQAQOQIMLTRISRS-----ESTVAMFKDERIESAGRYLSSDKKSNHK 1574
Qy      481 VSVLTVLHODWLNKGEYKCKVSNKALPAPIETKTSKANGQPREPOVYTLPPS----- 533
Db      1575 LV-CHAVQSQD--TGK-YRCVVTNKYGAESSECNVAV-----EDVTKFIAPFSATLSD 1624
Qy      534 RDLTLKNQVSLTLYVGFYSDIAVWESNGQENNYKTPPLVDSDGFLLYKSLTVDK 533
Db      1625 STALLGHNTLLECKVGSAPAE--VSWYDKGERISTTRIRQTDQNGN---CKLSISK 1678
Qy      594 SRWQGNVFSQS 605
Db      1679 AESDDMGVYVCS 1690

RESULT 98
Q8ISF4 PRELIMINARY; PRT; 2708 AA.
AC Q8ISF4;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, last annotation update)
DE 301kDa_1 protein.
GN ISOF.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Pelodertinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22269627; PubMed=12381307;
RA Flaherty D., Gernert K., Shmeleva N., Tang X., Mercer K.,
RA Borodovsky M., Benian G.;
RT "Titins in Caenorhabditis elegans with Unusual Features: Coiled-coil
RT Domains, Novel Regulation of Kinase Activity and Two New Possible
RT Elastic Regions."

```

RL J. Mol. Biol. 323:533-549 (2002).
 DR EMBL: AY130758; AAM61520.1; -.
 DR GO: GO:0005524; F:ATP binding; IEA.
 DR GO: GO:0004674; F:protein serine/threonine kinase activity; IEA.
 DR GO: GO:0004713; F:protein-tyrosine kinase activity; IEA.
 DR GO: GO:0004668; P:protein amino acid phosphorylation; IEA.
 DR InterPro: IPR003961; FN_III.
 DR InterPro: IPR007110; IG_1ike.
 DR InterPro: IPR003598; IG_c2.
 DR InterPro: IPR000719; Prot_kinase.
 DR InterPro: IPR002290; Ser_thr_kinase.
 DR InterPro: IPR008271; Ser_thr_pkin_AS.
 DR InterPro: IPR001245; Tyr_pkinase.
 DR Pfam: PF00041; fn3; 2.
 DR Pfam: PF00047; Ig; 19.
 DR Pfam: PF00069; pkinase; 1.
 DR ProDom: PD000001; Prot_kinase; 1.
 DR SMART: SM00408; ICG2; 19.
 DR SMART: SM00220; S_TKc; 1.
 DR SMART: SM00219; TyKc; 1.
 DR PROSITE: PS00835; IG_LIKE; 18.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS0011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
 DR SEQUENCE 2708 AA; 300729 MW; 28BEEB856FA571 CRC64;

Query Match 5.3%; Score 179.5; DB 5; Length 2708;
 Best Local Similarity 19.8%; Pred. No. 0.00034;
 Matches 133; Conservative 106; Mismatches 258; Indels 175; Gaps 32;
 QY 28 VILGKGDVVELTCTSSQKKSIOFHKNSQIKILNGQSFLTKGSKLNDRAEDRSRLM 87
 DB 1115 VILKTAGEATFTFQSYVAPAAQVW-----LHNGALQOTKSNYTRLF 1159
 QY 88 DQGNFLIKNLKLEISDFYICEVEDQKEV-----OLVGLTANS----- 129
 DB 1160 DQNTATLIVENVDDELGGITVANNQFGDVHTSAQITSGSBAKILASLPFIETELK 1219
 QY 130 DTHLLQGQSITLTLESPPGSSP-----SVQCRSP-----GKNIQGGKTLVSQ 173
 DB 1220 KINVEG--ATLISIQADLNGSPIDPEVWLKDNSELVESDRIQMKCGVNVQ-----LLVND 1273
 QY 174 LEIQDSGTCTCTYLQNGKKEFEKIDIVLAFOQASIVYKKEEQVEF-----S 222
 DB 1274 VGLDEGTITTTNENKGIKIQNTEVSV--TTSKVEKKEKKEKKEGKKRGKRG 1330
 QY 223 FPL--AFVTEKLT-----GSGELMQAERASSSKSWTF-DLNKEVSVK----- 264
 DB 1331 LPPPSGASKEQVTMAFDAPSEGPADSYEVERCPCPQRENVSGSTKSELEIKGLTPMT 1390
 QY 265 ----RVTDPEKLOMGKKLPLHLTLPOA---LPVAGSGNLTALBAKTKGLHQEVNLV 316
 DB 1391 EYIFRAVAGKNGKQGLGEMSEWSTLTKNASVQGAPOF-----TISPSK----- 1435
 QY 317 MRATLOKNTLTCVWGPSTSKMLSLKLENKAKVSKREK-----PWVNL-----PEA 365
 DB 1436 NRDEFE--IAVEFSGTPP-----SVKWKYKEMQIIVDEKIDVATTSTSSINLKQSE 1489
 QY 366 GMMQCLL-SPDSGCVLLESNIKVLPTWSTPVCPAPRPSKCDKHTCPBELLGSRVFLP 424
 DB 1490 GTNCLLENLQASASCOVTT--FNKPSLSTSTDSHSLEK-----LVP 1533
 QY 425 KPR----DTLMISRTPEVTCVVVDVSHEDPEVKFNNYVVDGVEVNAKTKRREQNSTYR 480
 DB 1534 LQKALNNESQAQOQIMLTCRISRS--ESTVAFKDERIESAGRYELSDKSKSNK 1589
 QY 481 VSVLVLYLHODVNLGKRYKRVSNKALPAPIETKTSKAKQPREPOVYTLPPS----- 533
 DB 1590 LV-CHAVQSD--TGK-YRCVVNKKYGAISECNVAV-----EDYTKFIAPSPSATLSD 1639
 QY 534 RDELTKNQSILCTLVGFGFYPSDI AVEWESNGQEPENNYKTPVLTLDGSGFFLTKSLTVDK 593

DB 1640 STAIGHNITLCEKVEGSPAPE--VSWTKDGERISTRRIRQTOPDENGN-----CKLSTSK 1693
 QY 594 SRMQGQNVFSGS 605
 DB 1694 AESDDMGVYVCS 1705

RESULT 99
 Q8MNSO PRELIMINARY; PRT; 2780 AA.
 AC Q8MNSO;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein.
 GN F12F3.2.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
 OC Rhabditidae; Pelodertinae; Caenorhabditis.
 OC NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=99069613; PubMed=9851916;
 RA Waterston R.;
 RT "Genome sequence of the nematode C. elegans: a platform for
 RT investigating biology. The C. elegans Sequencing Consortium.";
 RT Science 282:2012-2018 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Fulton B., Wohlmann P.;
 RT "The sequence of C. elegans cosmid F12F3.";
 RT submitted (JAN-1997) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Waterston R.;
 RT Submitted (May-2002) to the EMBL/Genbank/DBJ databases.
 RL EMBL: U80022; AAM29672.1; -.
 DR WormBep: F12F3.2a; CB30753.
 DR GO: GO:0005524; F:ATP binding; IEA.
 DR GO: GO:0004713; F:protein serine/threonine kinase activity; IEA.
 DR GO: GO:0004674; F:protein-tyrosine kinase activity; IEA.
 DR GO: GO:0016740; F:transferase activity; IEA.
 DR GO: GO:0004668; P:protein amino acid phosphorylation; IEA.
 DR InterPro: IPR003962; FNIII_subd.
 DR InterPro: IPR008957; FN_III-like.
 DR InterPro: IPR007110; IG_1like.
 DR InterPro: IPR003598; IG_c2.
 DR InterPro: IPR000719; Prot_kinase.
 DR InterPro: IPR002290; Ser_thr_kinase.
 DR InterPro: IPR008271; Ser_thr_pkin_AS.
 DR InterPro: IPR001245; Tyr_pkinase.
 DR Pfam: PF00041; fn3; 2.
 DR Pfam: PF00047; Ig; 19.
 DR Pfam: PF00069; pkinase; 1.
 DR PRINTS: PR00014; FNTYPEB11.
 DR ProDom: PD000001; Prot_kinase; 1.
 DR SMART: SM00408; ICG2; 19.
 DR SMART: SM00220; S_TKc; 1.
 DR SMART: SM00219; TyKc; 1.
 DR PROSITE: PS00835; IG_LIKE; 18.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS0011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
 KW Hypothetical protein; ATP-binding; Immunoglobulin domain; Repeat;
 KW Transferrase.
 SQ SEQUENCE 2780 AA; 308646 MW; 046D057107B935C1 CRC64;
 Query Match 5.3%; Score 179.5; DB 5; Length 2780;

QY	28	VVLGKGGDTVELTCTASQKKSIQFHWNKSNQIKLIGNQGSFLYTKGPSKLNDRADSRSLW	87
Db	1187	VVLKAGETATFTCSQSYANPAQVW-----LHNGKALQGTSKNKTBLF	1231
QY	88	DQGNPELLIKULKIEDPTIYCEVDQKEV-----QLLVFGLTANS-----	129
Db	1232	DDNATATLIEVNTBELCGTYAVANNQFGDVHTSAQLTISGSEAKKIAASLPYFIIEIKP	1291
QY	130	DTHLLOGSLTLTSPGSSP-----SVACRSR-----GKNIOGKTLVSQ	173
Db	1292	KINVEG--ATLSTADLNGSPIPEVWMLKNSLSEVSDRIOMKCDGVNYQ---LVYD	1345
QY	174	LELDGSGTWTCTVLQNKQKVEFKDIVLAFQKASSIYKKEGEVEF-----S	222
Db	1346	VGLEDEGTYYTTAENKGIHQNTSEVSV---TKSEVNEKEKKKKVKKEDGKKKRGPG	1400
QY	223	FPL--ATTVEKLT-----GSELMWQAEKASSSKSWITF-DLKNKESVK-----	264
Db	1403	LPRPSGAKTQVYTMFADPSPBGPADSYEVERRCPDQREWVSCGSTSLLEIKLTPTNT	1467
QY	265	---RVTDGPKLWQKDLPLHLTPQA---LPOYAGSGNLTALAEATGKLHGEVNLVY	316
Db	1463	EYIFRVAGKAKQGLGESEKMTSLTKTASVGAPOF-----TISPSK-----	1507
QY	317	MRATOLQKULTCENVGPTSPKMLSLKLENKAKYSKREK-----PWWVLN---DEA	365
Db	1508	NRDDEFE--IAVEPSGTPTP---SVKNYKEMQLQVPEBKIDVATSTSSILNLKSGEN	1561
QY	366	GMMQCLL-SDSGQVLLBSNIKVLPWSTPVPCCPAPBPKSCDKTHTCPBLGGPSVFLFP	424
Db	1562	GFNCLINENELGQASASCQVLT---FNKRPASTQSTPDSHLERN-----LVPT	1605
QY	425	KPK----DTLWISRTPEVTCVWVDVSHDDPEKFWVYVDGVEVHNAKTKKPREQYNSIYR	480
Db	1606	LOKALNNESAQAQOQIMLTCRISRS---ESTVAMFEDDERIESAGYELSSDKKSNHK	1661
QY	481	VSVVLTVLHODMLNGKEKCKVSNKALPAPIEKTISKAKGQPREQVYTLPPS-----	533
Db	1662	LV-CHAVGSDP--YCK-YRCVVTNKGVAESCNVAV-----EVTGFAPSPSATISD	1711
QY	534	RDELTKNOVSLTCLVKGFPSPDIAVWESNGQPENNYKTPPVLDSGSLYSLKLTVDK	593
Db	1712	STALIGHNITFLECKVEGSPAPB--VSMTKDGERISTTRIRIQTOENGN---CLSLSK	1765
QY	594	SRMOGQNVFSCS 605	
Db	1766	AESDDMGVYCS 1777	

RESULT 100

Q8MNS1 PRELIMINARY: PRT: 2808 AA.

AC Q8MNS1: 01-OCT-2002 (TREMBLrel. 22, Created)

DP 01-OCT-2002 (TREMBLrel. 22, Last sequence update)

DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)

GN Hypothetical protein.

OS *Caenorhabditis elegans*.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Pelodestinae; Caenorhabditia.

NCBI_TaxID=6239;

RA MEDLINE=99069613; PubMed=9851916;

RT "Genome sequence of the nematode *C. elegans*: a platform for

ST investigating biology. The *C. elegans* Sequencing Consortium. "

Science 282:2012-2018(1998).

RP	SEQUENCE FROM N.A.	[2]	
RC	STRAIN=Bristol N2;		
RA	Fulton B., Wohlmann P.;		
RT	"The sequence of <i>C. elegans</i> cosmid F12f3."		
RL	Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Bristol N2;		
RA	Waterston R.;		
RL	Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.		
DR	EMBL: U80022; AAM29673.1; -		
DR	WormRep; F12f3.2b; CE30754.		
DR	GO; GO:0005524; P:ATP binding; IEA.		
DR	GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.		
DR	GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.		
DR	GO:0016740; F:transferase activity; IEA.		
DR	GO:0006468; P:protein amino acid phosphorylation; IEA.		
DR	InterPro; IPR003962; PhII_subb.		
DR	InterPro; IPR003961; FN_III.		
DR	InterPro; IPR007110; IG_1like.		
DR	InterPro; IPR003598; IG_C2.		
DR	InterPro; IPR000719; Prot_kinase.		
DR	InterPro; IPR002290; Ser_Chk_kinase.		
DR	InterPro; IPR008271; Ser_Chk_pkin_AS.		
DR	InterPro; IPR01245; Tyr_pkinase.		
DR	Pfam; PF00041; fn3; 2.		
DR	Pfam; PF00047; Ig; 19.		
DR	Pfam; PF00069; pkinase; 1.		
DR	PRINTS; PR00014; FMYPRHII.		
DR	ProDom; PD000001; Proc_kinase; 1.		
DR	SMART; SM00060; FN3; 2.		
DR	SMART; SM00408; IGC2; 19.		
DR	SMART; SM00220; S_TKC; 1.		
DR	SMART; SM00219; Tyf_KC; 1.		
DR	PROSITE; PSS0835; IG_LIKE; 18.		
DR	PROSITE; PSS0107; PROTEIN_KINASE_ATP; 1.		
DR	PROSITE; PSS0011; PROTEIN_KINASE_DOM; 1.		
DR	PROSITE; PS0108; PROTEIN_KINASE_ST; 1.		
KW	Hypothetical protein; ATP-binding; Immunoglobulin domain; Repeat; Transferase.		
SO	SEQUENCE 2808 AA; 311979 MW; 02AFEDDAFE06FE12 CRC64;		
Query Match	5.3%; Score 179.5; DB 5; Length 2808;		
Best Local Similarity	19.8%; Pred. No. 0.00036;		
Matches 133; Conservative 106; Mismatches 258; Indels 175; Gaps 32;			
QY	28 VILGKKDDYELTCTASQKSKIQPHWKNQIKLNGQSFLLTKGPEKLANDRADRSKSLW 87		
DB	1187 VILKTAETATFTFCQSYANPAQVW-----LHNGKALQOTKSNVYKTRLF 1231		
QY	88 DQGNFPLIILNKLIKEDSDTYICEVEDQKEV-----OLVFGLTANS----- 129		
DB	1232 DNTATILVIEVDDELCGITTAVANNOFGDVHTSAQLTIGSEAKKIATASLPYFIIEIKP 1291		
QY	130 DTHLLQGSITLTLLESPGSSP-----SVQCRSPR-----GKNIQGGKTLISVSQ 173		
DB	1292 KINWVEG--ATLSIGADLNGSPPIPEVYWLKDNSELVESDRILQMKCGVNNQ---LLVRD 1345		
QY	174 LELDQSGTWCTYVLQNKQKVEPKIDIVLAFQKASSIVYKKEGQVEF-----S 222		
DB	1346 VGLDEDSCTYITTAENEKGIROWTEVASV---TSKVEYKKEKKKVKKEKDGKKKPGRPG 1402		
QY	223 FPL---AFVTEKLT-----GSGELMMAQFASSSKSWITF-DLKNKEVSVK----- 264		
DB	1403 LPRPSSGSKTEQVYMAFADSEGDADSEVERRCPPDRENVSSCGSTSLLELTIGLTPNT 1462		
QY	265 ----RVYQDRLQWKKLPLHLTPQK---LPQYASGNLTALAEKTEKLHQEVNLVV 316		
DB	1463 EYIFRVAGKKKQGGEGSEWSTSLTKTASVGQAQF-----TISPSQK-----IIA 1507		
QY	317 MRATQLOKNLTCVEWGFSTSPCLMLSLKLENKEAKVSKREK-----PVAVLN---PEA 365		

```

Db 1508 NRDEFE--IAVEFSGTPTP---SVKWKYKENVQIVPEKIDVATSTSSILNLKSQEN 1561
QY 366 GMMQCLL-SDSGVLLIESNIKVLPWTSTVPCCPAPBPSCDKHTHCPELLGGPSVFLPP 424
Db 1562 GTFNCLIEENELGQASASCQVTI---FNKPASLOSTPDHSLERN-----LVPT 1605
QY 425 KPK---DTLMTSRTEPVTCVVVDVSHEDPEVKFNNYVVDGVEVHNAKTKPREEQYNSTYR 480
Db 1606 LQKALNBSAQAGQOIMLTGRISRS---ESTVAMFKDDERIESAGRYELSSDKKSNHK 1661
QY 481 VVSVLTVLHODWMLNGKEYKKVSNKALPAPIEKTISKAKGQPREPQVYTLPPS----- 533
Db 1662 LV-CHAVQSD--TGK-YRCVVTNKKYGAISECNVAV-----EDVTKFIAPSFATLSD 1711
QY 534 RDELTKQVSLTCLVKGFPSPDIIVEMESNGQPENNYKTTTPVLDSGSPFLYSKLTVDK 593
Db 1712 STAILGHNITLLECKVEGSPAPE--VSMYKDGERTSTTRIRIOTODENGN---CKLISK 1765
QY 594 SRWQGNVFCSS 605
Db 1766 AESDDMGVYCS 1777

```

Search completed: August 3, 2004, 13:13:49
 Job time : 53.4955 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 3, 2004, 12:50:44 ; Search time 8.8991 Seconds
(without alignments)
3791.557 Million cell updates/sec

Title: SEQ07
Perfect score: 3414
Sequence: 1 MNRGVPPRHLLVQLALP.....DETCAQAQDELGLWTTDP 648

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 125 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2015	59.0	458	1	CD4_HUMAN
2	1991	58.3	458	1	CD4_PANTR
3	1844	54.0	458	1	CD4_MACFU
4	1843	54.0	458	1	CD4_MACMU
5	1840	53.9	458	1	CD4_MACPA
6	1835	53.7	458	1	CD4_MACNE
7	1818	53.3	458	1	CD4_CERAE
8	1717	50.3	397	1	CD4_BRYPA
9	1707	50.0	397	1	CD4_CERTO
10	1572.5	46.1	457	1	CD4_SALSC
11	1211.5	35.5	330	1	GC1_HUMAN
12	1150.5	33.7	459	1	CD4_RABIT
13	1143	33.5	290	1	CD4_HUMAN
14	1135	33.2	463	1	CD4_CANPA
15	1115.5	32.7	326	1	GC2_HUMAN
16	1105.5	32.4	327	1	GC4_HUMAN
17	1000.5	29.3	457	1	CD4_RAT
18	993.5	29.1	457	1	CD4_MOUSE
19	938.5	27.5	398	1	GC3_MOUSE
20	928.5	27.2	393	1	GC1W_MOUSE
21	910	26.7	323	1	GC1_RABIT
22	902.5	26.4	399	1	GCAM_MOUSE
23	879.5	25.8	329	1	GC2_MOUSE
24	879.5	25.8	405	1	GCBM_MOUSE
25	840.5	24.6	329	1	GC3_MOUSE
26	829	24.3	326	1	GC1_RAT
27	826.5	24.2	333	1	GCB_RAT
28	822.5	24.1	329	1	GC1_RAT
29	814.5	23.9	324	1	GC1_MOUSE
30	801.5	23.5	330	1	GCAM_MOUSE
31	792	23.2	332	1	GC1_RAT
32	788	23.1	335	1	GCAB_MOUSE
33	783.5	22.9	336	1	GCB_MOUSE

34	368.5	10.8	421	1	EPC_MOUSE
35	363.5	10.6	429	1	EPC_RAT
36	359.5	10.5	476	1	MUC_MOUSE
37	358.5	10.5	454	1	MUC_HUMAN
38	356	10.4	428	1	EPC_HUMAN
39	355.5	10.4	479	1	MUC_RABIT
40	352.5	10.3	391	1	MUCB_HUMAN
41	348	10.2	455	1	MUC_MOUSE
42	346	10.1	457	1	MUC_GUNNU
43	346	10.1	458	1	MUC_RABIT
44	342.5	10.0	454	1	MUC_MESAU
45	340	10.0	450	1	MUC_CANPA
46	306	9.0	299	1	ALC_RABIT
47	303	8.9	461	1	HVCM_HETER
48	301.5	8.8	438	1	HVCS_HETER
49	300	8.8	446	1	MUC_CHICK
50	297.5	8.7	438	1	HVC2_HETER
51	293.5	8.6	340	1	ALC2_HUMAN
52	289.5	8.5	393	1	HVC3_HETER
53	282.5	8.3	353	1	ALC1_GORGO
54	282.5	8.3	353	1	ALC1_HUMAN
55	270	7.9	370	1	HVCM_HETER
56	261.5	7.7	481	1	MUCM_ICTPU
57	251.5	7.4	344	1	ALC_MOUSE
58	199.5	5.8	513	1	SHS1_MOUSE
59	184	5.4	4391	1	PGBM_HUMAN
60	177.5	5.2	105	1	LAC1_MOUSE
61	177.5	5.2	3707	1	PGBM_MOUSE
62	174.5	5.1	739	1	VCA1_RAT
63	174	5.1	6632	1	UN89_CAEEL
64	172	5.0	104	1	LAC2_MOUSE
65	171	5.0	506	1	SHS1_BOVIN
66	165	4.8	702	1	CEAS_HUMAN
67	163	4.8	104	1	LAC3_MOUSE
68	162.5	4.8	213	1	ILL1_HUMAN
69	162.5	4.8	1493	1	NEOL_MOUSE
70	162	4.7	1906	1	KML5_CHICK
71	161	4.7	104	1	LAC2_MOUSE
72	161	4.7	268	1	HB2D_PIG
73	160.5	4.7	739	1	VCA1_HUMAN
74	159	4.6	105	1	LAC_HUMAN
75	156	4.6	105	1	LAC_PIG
76	153.5	4.5	105	1	LAC5_MUSSP
77	153	4.5	105	1	LAC_RABIT
78	153	4.5	106	1	KACB_RABIT
79	152.5	4.5	847	1	CD22_HUMAN
80	151.5	4.4	383	1	DTG_HUMAN
81	150	4.4	106	1	KACB_RAT
82	149.5	4.4	1447	1	DCC_MOUSE
83	148.5	4.3	1257	1	CAML_HUMAN
84	148	4.3	106	1	KAC_HUMAN
85	147	4.3	739	1	VCA1_MOUSE
86	146	4.3	104	1	LAC1_RAT
87	146	4.3	387	1	SRB2_HUMAN
88	145.5	4.3	103	1	LAC_CHICK
89	145	4.2	106	1	KACB_RAT
90	145	4.2	261	1	HB2C_PIG
91	145	4.2	509	1	SHS1_RAT
92	145	4.2	837	1	NCM2_MOUSE
93	144.5	4.2	555	1	C166_CARAU
94	143.5	4.2	503	1	SHS1_HUMAN
95	143.5	4.2	1260	1	SHS1_MOUSE
96	143	4.2	106	1	KAC_MOUSE
97	142.5	4.2	398	1	SRB1_HUMAN
98	141.5	4.1	1447	1	DCC_HUMAN
99	141	4.1	6885	1	SNE2_HUMAN
100	140	4.1	268	1	HB2X_HUMAN
101	140	4.1	273	1	ZDOB_HUMAN
102	140	4.1	1259	1	CAML_RAT
103	139.5	4.1	2012	1	DSOB_HUMAN
104	139	4.1	273	1	CD4_PANTR
105	138.5	4.1	564	1	C166_BRARE
106	138.5	4.1	1197	1	CAM1_BRARE

P06336	mus musculus
P01855	rattus norv
P01873	mus musculus
P01871	homo sapien
P01854	homo sapien
P04221	oryctolagus
P04220	homo sapien
P01872	mus musculus
P20768	suncus murti
P03988	oryctolagus
P06337	mesocricetu
P01874	canis famli
P01879	oryctolagus
P23088	heterodontu
P23087	heterodontu
P01875	gallus gall
P23085	heterodontu
P01877	homo sapien
P23086	heterodontu
P20758	gorilla gor
P01876	homo sapien
P23084	heterodontu
P23735	ictalurus p
P01878	mus musculus
P97797	m. protei-t
P98160	homo sapien
P01843	mus musculus
Q05793	mus musculus
P29534	rattus norv
O01761	caenorhabdi
P20767	rattus norv
O46631	bos taurus
P06731	homo sapien
P01845	mus musculus
P15814	homo sapien
P97798	mus musculus
P11799	gallus gall
P01844	mus musculus
P15980	sus scrofa
P19320	homo sapien
P01842	homo sapien
P01846	sus scrofa
P20765	mus apretus
P01847	oryctolagus
P01834	oryctolagus
P20278	homo sapien
P01880	homo sapien
P70211	mus musculus
P32004	homo sapien
P01834	homo sapien
P29533	mus musculus
P20766	rattus norv
O91976	rattus norv
P20763	gallus gall
P01835	rattus norv
P15982	sus scrofa
O35136	mus musculus
P01837	mus musculus
O00241	homo sapien
P33146	homo sapien
O8wx10	homo sapien
P05558	homo sapien
P13765	homo sapien
Q05695	rattus norv
O60469	homo sapien
P18467	pan troglod
O90460	brachydanio
Q90478	brachydanio

```

107 139 4.0 997 1 SCPI RAT 003410 rattus norv
108 137.5 4.0 761 1 NCAG_HUMAN P13592 homo sapien
109 136.5 4.0 1348 1 VGR2_COTJA P52583 coturnix co
110 136 4.0 261 1 HB24_HUMAN P01920 homo sapien
111 135.5 4.0 105 1 LACS_MOUSE P20764 mus musculu
112 134.5 3.9 261 1 HB22_HUMAN P01919 homo sapien
113 134 3.9 1369 1 NFAS_CHICK 042414 gallus gall
114 133 3.9 837 1 NCW2_HUMAN 015392 homo sapien
115 132.5 3.9 261 1 HB21_HUMAN P01918 homo sapien
116 132.5 3.9 1240 1 NFAS_MOUSE Q81003 mus musculu
117 132 3.9 231 1 HB21_CHICK P23068 gallus gall
118 131.5 3.9 848 1 NCAL_HUMAN P13591 homo sapien
119 131 3.8 261 1 HB23_HUMAN P05537 homo sapien
120 130 3.8 1277 1 CAML_FUGRU P20241 drosophila
121 130 3.8 1302 1 NRG_DROME P24969 haemophilus
122 130 3.8 1694 1 TGAO_HAEIN P42782 haemophilus
123 129 3.8 1541 1 IGAL_HAEIN P97685 rattus norv
124 128.5 3.8 1240 1 NFAS_RAT P97685 rattus norv
125 128 3.7 1141 1 MYPS_HUMAN Q00872 homo sapien

```

ALIGNMENTS

```

RESULT 1
ID_CD4_HUMAN STANDARD; PRT; 458 AA.
AC P01730;

```

```

DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
DE T4/Len-3).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85254948; PubMed=2990730;
RA Maddon P.J., Littman D.R., Godfrey M., Maddon D.E., Chess L.,
RA Axel R.;
RT "The isolation and nucleotide sequence of a cDNA encoding the T cell
RT surface protein T4: a new member of the immunoglobulin gene family.";
RL Cell 42:93-104(1985).
RN [2]
RP REVISION TO 26.
RX MEDLINE=89028665; PubMed=3263213;
RA Littman D.R., Maddon P.J., Axel R.;
RT "Corrected CD4 sequence.";
RL Cell 55:541-541(1988).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=96303695; PubMed=8723724;
RA Anasetti-Lari M.A., Muzny D.M., Lu J., Lu F., Lilly C.E., Spanos S.,
RA Malley T., Gibbs R.A.;
RT "A gene-rich cluster between the CD4 and triosephosphate isomerase
RT genes at human chromosome 12p13.";
RL Genome Res. 6:314-326(1996).
RN [4]
RP SEQUENCE FROM N.A. AND VARIANT TRP-265.
RX MEDLINE=91216786; PubMed=1708753;
RA Hodge T.W., Sasso D.R., McDougal J.S.;
RT "Humans with OKT4-epitope deficiency have a single nucleotide base
RT change in the CD4 gene, resulting in substitution of TRP240 for
RT ARG240.";
RL Hum. Immunol. 30:99-104(1991).
RN [5]
RP SEQUENCE FROM N.A.
RX TISSUE=Pancreas;
RC MEDLINE=22388257; PubMed=12477932;
RA Straubeberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

```

```

RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heish F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M.J., Soares M.B., Bonaldo M.F., Caesvanti P., Scheetz T.E.,
RA Brownstein M.J., Uedlin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mallary S.J.,
RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman W., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green B.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Buterfield V.S.N., Krzywicki M.I., Skalska U., Smalls D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Maitra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [6]
RP SEQUENCE OF 28-424 FROM N.A.
RC TISSUE=Blood;
RX MEDLINE=93049640; PubMed=1425921;
RA Fomsgaard A., Hirsch V.M., Johnson P.R.;
RT "Cloning and sequences of primate CD4 molecules: diversity of the
RT cellular receptor for simian immunodeficiency virus/human
RT immunodeficiency virus.";
RL Eur. J. Immunol. 22:2973-2981(1992).
RN [7]
RP SEQUENCE OF 26-394.
RX MEDLINE=90078232; PubMed=2592374;
RA Carr S.A., Hemling M.E., Folea-Wasserman G., Sweet R.W., Anumula K.,
RA Barr J.R., Huddleston M.J., Taylor P.;
RT "Protein and carbohydrate structural analysis of a recombinant
RT soluble CD4 receptor by mass spectrometry.";
RL J. Biol. Chem. 264:21286-21295(1989).
RN [8]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 26-208.
RX MEDLINE=91061881; PubMed=1701030;
RA Wang J., Yan Y., Garrett T.P., Liu J., Rodgers D.W., Garlick R.L.,
RA Tarr G.E., Husein Y., Reinherz E.L., Harrison S.C.;
RT "Atomic structure of a fragment of human CD4 containing two
RT immunoglobulin-like domains.";
RL Nature 348:411-418(1990).
RN [9]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 26-208.
RX MEDLINE=91061882; PubMed=2247146;
RA Ryl S.-E., Kwong P.D., Truneh A., Porter T.G., Arthos J.,
RA Rosenberg M., Dai X., Xuong N.-H., Axel R., Sweet R.W.,
RA Hendrickson W.A.;
RT "Crystal structure of an HIV-binding recombinant fragment of human
RT CD4.";
RL Nature 348:419-426(1990).
RN [10]
RP X-RAY CRYSTALLOGRAPHY (3.9 ANGSTROMS) OF 26-388.
RX MEDLINE=97311402; PubMed=9168119;
RA Wu H., Kwong P.D., Hendrickson W.A.;
RT "Dimeric association and segmental variability in the structure of
RT human CD4.";
RL Nature 387:527-530(1997).
RN [11]
RP PALMITOYLATION.
RX MEDLINE=92317088; PubMed=1618861;
RA Criste B., Rose J.K.;
RT "Identification of palmitoylation sites on CD4, the human
RT immunodeficiency virus receptor.";
RL J. Biol. Chem. 267:13593-13597(1992).
RN [12]
RP FUNCTION: Accessory protein for MHC class-II antigen/T-cell
CC receptor interaction. May regulate T-cell activation.
CC -I- SUBUNIT: Associates with p56-lck.
CC -I- SUBCELLULAR LOCATION: Type I membrane protein.
CC -I- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -I- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.

```

```

CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD4 entry;
CC WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd4.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; M12807; AAA5572.1; -
DR EMBL; U47924; AAB51309.1; -
DR EMBL; M35160; AAA16069.1; -
DR EMBL; BC025782; AAB25782.1; -
DR PIR; A90872; RHMUT4.
DR PDB; 1CDH; 30-APR-94.
DR PDB; 1CDI; 30-APR-94.
DR PDB; 3CD4; 31-OCT-93.
DR PDB; 1CDJ; 01-APR-97.
DR PDB; 1CDU; 01-APR-97.
DR PDB; 1CDB; 01-APR-97.
DR PDB; 1CDB; 12-MAR-97.
DR PDB; 1WIP; 07-JUL-97.
DR PDB; 1WIP; 07-JUL-97.
DR PDB; 1WIP; 07-JUL-97.
DR PDB; 1WIP; 07-JUL-97.
DR PDB; 1G9M; 27-DEC-00.
DR PDB; 1G9N; 27-DEC-00.
DR PDB; 1GCI; 19-AUG-98.
DR PDB; 1JL4; 19-SEP-01.
DR GlycositeDB; P01730; -
DR GlycositeDB; P01730; -
DR HGN; 186940; -
DR GO; GO:0042101; C:T-cell receptor complex; NAS.
DR GO; GO:0015026; F:coreceptor activity; NAS.
DR GO; GO:0015029; F:internalization receptor activity; NAS.
DR GO; GO:0042289; F:MHC class II protein binding; NAS.
DR GO; GO:0004888; F:transmembrane receptor activity; NAS.
DR GO; GO:0006955; P:immune response; NAS.
DR GO; GO:0009405; P:pathogenesis; NAS.
DR GO; GO:0045086; P:positive regulation of interleukin-2 biosyn. . . NAS.
DR GO; GO:0030217; P:T-cell differentiation; NAS.
DR GO; GO:0045058; P:T-cell selection; NAS.
DR GO; GO:0007169; P:transmembrane receptor protein tyrosine kin. . . NAS.
DR InterPro; IPR000973; CD4_TcAg.
DR InterPro; IPR007110; Ig_Like.
DR Pfam; PF00047; Ig_2.
DR PRINTS; PRO0692; CD4TCANTIGEN.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_Like; 1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Immune response; Repeat; Signal; Lipoprotein; Palmitate; 3D-structure;
KW Polymorphism.
FT SIGNAL 1 25
FT CHAIN 26 458
FT DOMAIN 26 396
FT TRANSMEM 397 418
FT DOMAIN 419 458
FT DOMAIN 26 125
FT DOMAIN 126 203
FT DOMAIN 204 317
FT DOMAIN 318 374
FT CARBOHYD 296 296
FT CARBOHYD 325 325
FT DISULFID 41 109
FT DISULFID 155 184
FT DISULFID 328 370
FT LIPID 419 419
FT LIPID 422 422
FT VARIANT 265 265
R -> W (in OKT4-negative populations).

```

```

FT STRAND 27 32 /FTID=VAR_003906.
FT TURN 33 34
FT STRAND 37 39
FT TURN 44 45
FT STRAND 51 55
FT TURN 56 57

Query Match 59.0%; Score 2015; DB 1; Length 458;
Best Local Similarity 99.5%; Pred. No. 3,4e-121;
Matches 392; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 MNRGVPFPHLLVQLALPAATGQKVVYKGGDTVELTCTASQKSIQFHMNKNQIK 60
Db 1 MNRGVPFPHLLVQLALPAATGQKVVYKGGDTVELTCTASQKSIQFHMNKNQIK 60
Oy 61 ILGNQGSFLTKGPKSKNDNRADSRSLWDQGNFPLIKNLKIEDSDTYICEVEDQKEVOL 120
Db 61 ILGNQGSFLTKGPKSKNDNRADSRSLWDQGNFPLIKNLKIEDSDTYICEVEDQKEVOL 120
Oy 121 LVFGLTANSTHLLQGSLLTLTLESPGSSPVQCRSPRGKNTGGKTLISVQLELDSG 180
Db 121 LVFGLTANSTHLLQGSLLTLTLESPGSSPVQCRSPRGKNTGGKTLISVQLELDSG 180
Oy 181 TWCTVQONQKVEFKIDIVLAFOKASSIVYKKEGQVEFSPPLATVEKLTGSGELMW 240
Db 181 TWCTVQONQKVEFKIDIVLAFOKASSIVYKKEGQVEFSPPLATVEKLTGSGELMW 240
Oy 241 QABRASSKSWITFDLKNKEVSRYVTDPRKQWKKPLHLTLPOALPOYAGSGNLTLA 300
Db 241 QABRASSKSWITFDLKNKEVSRYVTDPRKQWKKPLHLTLPOALPOYAGSGNLTLA 300
Oy 301 LEAKTGKHOEVLVYWRATQOLNLTCEWGPSPPLMLSLKLENKAVSKREKVVW 360
Db 301 LEAKTGKHOEVLVYWRATQOLNLTCEWGPSPPLMLSLKLENKAVSKREKVVW 360
Oy 361 LNPBAGWQCLLSDSGVLLSNKVLPTWSTVP 394
Db 361 LNPBAGWQCLLSDSGVLLSNKVLPTWSTVP 394

RESULT 2
CD4_PANTR STANDARD; PRT; 458 AA.
AC P16004;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
DE T4/Leu-3).
GN CD4.
OS Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
OX NCBI_TaxID=9598;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=9018264; Pubmed=2107024;
RX Camerini D., Seed B.;
RT "A CD4 domain important for HIV-mediated syncytium formation lies
RT outside the virus binding site.";
RL Cell 60:747-754(1990).
RN [2]
RP SEQUENCE OF 26-424 FROM N.A.
RA Fomsgaard A., Hirsch V.M., Johnson P.R.;
RX MEDLINE=93049640; Pubmed=1425921;
RT "Cloning and sequences of primate CD4 molecules: diversity of the
RT cellular receptor for simian immunodeficiency virus/human
RT immunodeficiency virus.";
RL Eur. J. Immunol. 22:2973-2981(1992).
CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
receptor interaction. May regulate T-cell activation.

```

```

CC -1- SUBUNIT: Associates with p56-lck (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M31135; AAA35407.1; -.
DR EMBL; X73323; CAAS1749.1; -.
DR PIR; B32722; RMC2T4.
DR HSSP; P01730; 1WIO.
DR GO; GO:0042101; C:T-cell receptor complex; ISS.
DR GO; GO:0015026; F:coreceptor activity; ISS.
DR GO; GO:0042289; F:MHC class II protein binding; ISS.
DR GO; GO:0006955; P:immune response; ISS.
DR GO; GO:0045086; P:positive regulation of interleukin-2 biosyn. . .; ISS.
DR GO; GO:0030217; P:T-cell differentiation; ISS.
DR GO; GO:0045058; P:T-cell selection; ISS.
DR GO; GO:0007169; P:transmembrane receptor protein tyrosine kin. . .; ISS.
DR InterPro; IPR000973; CD4_TCSAg.
DR InterPro; IPR007110; Ig_Like.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; Ig; 2.
DR PRINTS; PR00692; CD4TCANTIGEN.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_Like; 1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Immune response; Repeat; Signal; Lipoprotein; Palmitate.
FT SIGNAL 1 25
FT CHAIN 26 458
FT DOMAIN 26 396
FT TRANSMEM 397 418
FT DOMAIN 419 458
FT DOMAIN 26 125
FT DOMAIN 126 203
FT DOMAIN 204 317
FT DOMAIN 318 374
FT CARBOHYD 296 296
FT CARBOHYD 325 325
FT DISULFID 41 109
FT DISULFID 155 184
FT DISULFID 328 370
FT LIPID 419 419
FT LIPID 422 422
FT CONFLICT 42 42
FT CONFLICT 62 62
FT CONFLICT 191 191
SQ SEQUENCE 458 AA; 51057 MW; A7C3AC8A5257D3AD CRC64;

Query Match 58.3%; Score 1991; DB 1; Length 458;
Best Local Similarity 98.5%; Pred. No. 1.2e-119;
Matches 388; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

```

Db 181 TWTCVTQVONQKVEFKIDIVLAFQKASIVYKKEGEVFSFPLAFTVEKLTGSGELMW 240
Qy 241 QAEPRSSSSSWITPDLKKEVSVKVTODPKIQMGKPLPHLTLPALQVYAGSSNLTIA 300
Db 241 QAEPRSSSSSWITPDLKKEVSVKVTODPKIQMGKPLPHLTLPALQVYAGSSNLTIA 300
Qy 301 LEAKYKHLHQEYNLVVMRATQLOKNU/TCCEWGPSPKMLSLKLEKAKVSKEKPVWY 360
Db 301 LEAKYKHLHQEYNLVVMRATQLOKNU/TCCEWGPSPKMLSLKLEKAKVSKEKPVWY 360
Qy 361 LNPEAGMQLLSDSGQVLLBSNINVLPTWSPV 394
Db 361 LNPEAGMQLLSDSGQVLLBSNINVLPTWSPV 394

RESULT 3
ID CD4_MACFU STANDARD; PRT; 458 AA.
AC P79184;
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
DE T4/Les-3).
GN CD4.
OS Macaca fuscata fuscata (Japanese macaque).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoinae; Macaca.
OC NCBI_TaxID=9543;
RN [1]
RP SEQUENCE FROM N.A.
RA Hashimoto O., Tatsumi M.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
CC receptor interaction. May regulate T-cell activation.
CC -1- SUBUNIT: Associates with p56-lck (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D63348; BAA09672.1; -.
DR HSSP; P01730; 1WBR.
DR GO; GO:0042101; C:T-cell receptor complex; ISS.
DR GO; GO:0015026; F:coreceptor activity; ISS.
DR GO; GO:0042289; F:MHC class II protein binding; ISS.
DR GO; GO:0006955; P:immune response; ISS.
DR GO; GO:0045086; P:positive regulation of interleukin-2 biosyn. . .; ISS.
DR GO; GO:0030217; P:T-cell differentiation; ISS.
DR GO; GO:0045058; P:T-cell selection; ISS.
DR GO; GO:0007169; P:transmembrane receptor protein tyrosine kin. . .; ISS.
DR InterPro; IPR000973; CD4_TCSAg.
DR InterPro; IPR007110; Ig_Like.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; Ig; 2.
DR PRINTS; PR00692; CD4TCANTIGEN.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_Like; 1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Immune response; Repeat; Signal; Lipoprotein; Palmitate.
FT SIGNAL 1 25
FT CHAIN 26 458
FT DOMAIN 26 396
FT TRANSMEM 397 418
FT DOMAIN 419 458
SQ SEQUENCE 458 AA; 51057 MW; A7C3AC8A5257D3AD CRC64;

```

FT DOMAIN 26 125 IG-LIKE V-TYPE.
 FT DOMAIN 126 203 IG-LIKE C2-TYPE 1.
 FT DOMAIN 317 317 IG-LIKE C2-TYPE 2.
 FT DOMAIN 318 374 IG-LIKE C2-TYPE 3.
 FT CARBOHYD 42 42 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 296 296 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
 FT CARBOHYD 325 325 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
 FT DISULFID 41 109 BY SIMILARITY.
 FT DISULFID 155 184 BY SIMILARITY.
 FT DISULFID 328 370 BY SIMILARITY.
 FT LIPID 419 419 S-palmitoyl cysteine (By similarity).
 FT LIPID 422 422 S-palmitoyl cysteine (By similarity).
 SQ SEQUENCE 458 AA; 50828 MM; 76B3E7EF08185535 CRC64;

Query Match 54.0%; Score 1844; DB 1; Length 458;
 Beel Local Similarity 90.4%; Pred. No. 2.7e-110;
 Matches 356; Conservative 17; Mismatches 21; Indels 0; Gaps 0;

QY 1 MNRGVPFRLHLLVQLALLPAATOGNKKVLLGKGGDTVELTCTASOKKSIOFHMKNNOIK 60
 DB 1 MNRGVPFRLHLLVQLALLPAATOGNKKVLLGKGGDTVELTCTASOKKSIOFHMKNNOIK 60
 QY 61 ILGNQGSFLTKGPKSLNDRADSRSLMDQGNFPLIKKIEDSDTYICEVEDQKEEYOL 120
 DB 61 ILGNQGSFLTKGPKSLNDRADSRSLMDQGNFPLIKKIEDSDTYICEVEDQKEEYOL 120
 QY 121 LVFGLTANSDPHLLQGGSLTTLTSSPPSSPVOCRSRGNKIQGGKTLVSQHELDQSG 180
 DB 121 LVFGLTANSDPHLLQGGSLTTLTSSPPSSPVOCRSRGNKIQGGKTLVSQHELDQSG 180
 QY 121 LVFGLTANSDPHLLQGGSLTTLTSSPPSSPVOCRSRGNKIQGGKTLVSQHELDQSG 180
 DB 181 TWCTCTVQNOKKVEFKIDIVLAFQKASSIVYKKEGEVSPFLAFTVELTSGGELMW 240
 DB 181 TWCTCTVQNOKKVEFKIDIVLAFQKASSIVYKKEGEVSPFLAFTVELTSGGELMW 240
 QY 241 QABRASSKSWITFDLKNKEVSVKRVTDPRKQNGKKPLHLTLPOALPOYAGSGNTLA 300
 DB 241 QABRASSKSWITFDLKNKEVSVKRVTDPRKQNGKKPLHLTLPOALPOYAGSGNTLA 300
 QY 241 QABRASSKSWITFDLKNKEVSVKRVTDPRKQNGKKPLHLTLPOALPOYAGSGNTLA 300
 DB 241 QABRASSKSWITFDLKNKEVSVKRVTDPRKQNGKKPLHLTLPOALPOYAGSGNTLA 300
 QY 301 LEATGKLGHOEVNLVMAATQLOKRLTEWNGPSPKMLSLKLENEKAKSKREKPVW 360
 DB 301 LEATGKLGHOEVNLVMAATQLOKRLTEWNGPSPKMLSLKLENEKAKSKREKPVW 360
 QY 301 LEATGKLGHOEVNLVMAATQLOKRLTEWNGPSPKMLSLKLENEKAKSKREKPVW 360
 DB 301 LEATGKLGHOEVNLVMAATQLOKRLTEWNGPSPKMLSLKLENEKAKSKREKPVW 360
 QY 361 LNPEAGMWCILLSDSGOVLLESNIKVLPTWSTPV 394
 DB 361 LNPEAGMWCILLSDSGOVLLESNIKVLPTWSTPV 394

RESULT 4
 CD4_MACMU STANDARD; PRT: 458 AA.
 ID CD4_MACMU
 AC P16003; Q29617; STANFORD; PRT: 458 AA.
 DT 01-APR-1990 (Rel. 14, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen T4/Leu-3).
 GN CD4.
 OS Macaca mulatta (Rhesus macaque).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea; Macaca.
 OC NCBI_Taxid=9544;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90182664; PubMed=2107024;
 RA Cameron D., Seed B.;
 RT "A CD4 domain important for HIV-mediated syncytium formation lies outside the virus binding site";
 RL Cell 60:747-754 (1990).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC TISSUE=Thymocytes;
 RA Hashimoto O., Tateuchi M.;

RT "Molecular cloning and expression of macaque CD4s.";
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 28-424 FROM N.A.
 RC TISSUE=Blood;
 RX MEDLINE=93049640; PubMed=1425921;
 RA Fomsgaard A., Hirsch V.M., Johnson P.R.;
 RT "Cloning and sequences of primate CD4 molecules: diversity of the cellular receptor for simian immunodeficiency virus/human immunodeficiency virus";
 RL Eur. J. Immunol. 22:2973-2981 (1992).
 RN [4]
 RP SEQUENCE OF 107-192 FROM N.A.
 RX MEDLINE=98320644; PubMed=9656488;
 RA Harris E.E., Disotell T.R.;
 RT "Nuclear gene trees and the phylogenetic relationships of the mangabeys (Primates: Papionini).";
 RL Mol. Biol. Evol. 15:892-900 (1998).
 CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell receptor interaction. May regulate T-cell activation.
 CC -1- SUBUNIT: Associates with p56-lck (by similarity).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M31134; AAA36838.1; -;
 DR EMBL; D63347; BAA09671.1; -;
 DR EMBL; X73326; CAA51752.1; -;
 DR EMBL; AF057385; AAC25129.1; -;
 DR HSSP; P01730; 1MBR.
 DR GO; GO:0042101; C: T-cell receptor complex; ISS.
 DR GO; GO:0015026; F: coreceptor activity; ISS.
 DR GO; GO:0042289; F: MHC class II protein binding; ISS.
 DR GO; GO:0006955; P: immune response; ISS.
 DR GO; GO:0045086; P: positive regulation of interleukin-2 biosyn. . .; ISS.
 DR GO; GO:0030217; P: T-cell differentiation; ISS.
 DR GO; GO:0045058; P: T-cell selection; ISS.
 DR InterPro; IPR000973; CD4_TcAg.
 DR InterPro; IPR007110; Ig_Like.
 DR InterPro; IPR003596; Ig_V.
 DR Pfam; PF00047; Ig_2.
 DR PRINTS; PR00692; CD4TCANTIGEN.
 DR SMART; SM00406; IG; 1.
 DR PROSITE; PS50835; IG_LIKE; 1.
 KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell; Immune response; Repeat; Signal; Lipoprotein; Palmitate.
 FT SIGNAL 1 25
 FT CHAIN 26 458
 FT DOMAIN 26 396
 FT TRANSMEM 397 418
 FT DOMAIN 419 458
 FT DOMAIN 26 125
 FT DOMAIN 126 203
 FT DOMAIN 204 317
 FT DOMAIN 318 374
 FT CARBOHYD 296 296
 FT CARBOHYD 325 325
 FT DISULFID 41 109
 FT DISULFID 155 184
 FT DISULFID 328 370
 FT LIPID 419 419
 FT LIPID 422 422
 FT LIPID 422 422
 FT CONFLICT 62 62
 L -> S (IN REF. 3).

```

FT CONFLICT 67 67 L -> S (IN REF. 2).
FT CONFLICT 169 169 I -> L (IN REF. 2).
FT CONFLICT 191 191 K -> N (IN REF. 3).
FT CONFLICT 248 248 S -> P (IN REF. 2).
FT CONFLICT 265 265 R -> Q (IN REF. 3).
FT CONFLICT 349 349 A -> T (IN REF. 2).
SQ SEQUENCE 458 AA; 50884 MW; 8BB80339FAFEC808 CRC64;

Query Match 54.0%; Score 1843; DB 1; Length 458;
Best Local Similarity 90.4%; Pred. No. 3,1e-110;
Matches 356; Conservative 17; Mismatches 21; Indels 0; Gaps 0;

QY 1 MNRGVPFRHLVLVQLALPAATQGNKVVYGGKGDVETLCTASOKKSIOFHKNKSNQIK 60
DB 1 MNRGIFPRHLVLVQLALPAATQGNKVVYGGKGDVETLCTASOKKSIOFHKNKSNQIK 60
QY 1LGNQGSFLTKGSPSKINDRADSRSLMDQGNFPLIINKLIEDSDTYICEVEDQKEVOL 120
DB 61 ILGIQGFLLTKGSPSKINDRADSRSLMDQGSFMIINKLIEDSDTYICEVENKKEVEL 120
QY 121 LVFGLTANSDTHLLQGSQSLTLTLSPGSSPSVQCRSPGKNIOGKTLVSQLELQDSG 180
DB 121 LVFGLTANSDTHLLQGSQSLTLTLSPGSSPSVQCRSPGKNIOGKTLVSQLELQDSG 180
QY 181 TWTCTVLQNKQKVEFKIDIVLAFQKASSIVYKKEGQVEFSPPLAFTVEKLTGSGELMW 240
DB 181 TWTCTVSQDKQKVEFKIDIVLAFQKASSIVYKKEGQVEFSPPLAFTVEKLTGSGELMW 240
QY 241 QAEARASSKSWITFDLKNKEVSVKRYTQDPKLGKGLPLHLTLPALPOYAGSGNLTLLA 300
DB 241 QAEARASSKSWITFDLKNKEVSVKRYTQDPKLGKGLPLHLTLPALPOYAGSGNLTLLA 300
QY 301 LEAKTGKLEQEVNLVVMRATQLOKNTLCEVWGSPSPKMLSLKLENKAVSKREKPVVW 360
DB 301 LEAKTGKLEQEVNLVVMRATQLOKNTLCEVWGSPSPKMLSLKLENKAVSKREKPVVW 360
QY 361 LNPEAGMWOCCLSDSGOVLLESNIKVLPTWSTPV 394
DB 361 LNPEAGMWOCCLSDSGOVLLESNIKVLPTWSTPV 394

RESULT 5
CD4_MACPA STANDARD; PRT; 458 AA.
ID CD4_MACPA
AC P79185;
DC 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
DE T4/Leu-3).
GN Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Macaca.
OC NCBI_Taxid=9541;
RN [1]
RN SEQUENCE FROM N.A.
RC TISSUE=Thymocytes;
RA Tatsumi M., Yabe M., Yamada Y.K.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
CC receptor interaction. May regulate T-cell activation.
CC -1- SUBUNIT: Associates with p56-lck (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
EMBL; D63349; BAA09673.1; -.
DR HSSP; P01730; 1MR.
DR GO; GO:0042101; C1T-cell receptor complex; ISS.
DR GO; GO:0015026; F1coreceptor activity; ISS.
DR GO; GO:0042289; F1MHC class II protein binding; ISS.
DR GO; GO:0005955; P1immune response; ISS.
DR GO; GO:0045086; P1positive regulation of interleukin-2 biosyn. . .; ISS.
DR GO; GO:0030217; P1T-cell differentiation; ISS.
DR GO; GO:0045058; P1T-cell selection; ISS.
DR GO; GO:0007169; P1transmembrane receptor protein tyrosine kin. . .; ISS.
DR InterPro; IPR000973; CD4 TCR.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig_2.
DR PRINTS; PR00692; CD4TCANTIGEN.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Immune response; Repeat; Signal; Lipoprotein; Palmitate.
FT SIGNAL 1 25
FT CHAIN 26 458
FT DOMAIN 26 396
FT TRANSMEM 397 418
FT DOMAIN 419 458
FT DOMAIN 26 125
FT DOMAIN 126 203
FT DOMAIN 204 317
FT DOMAIN 318 374
FT CARBOHYD 42 42
FT CARBOHYD 296 296
FT CARBOHYD 325 325
FT DISULFID 41 109
FT DISULFID 155 184
FT DISULFID 328 370
FT LIPID 419 419
FT LIPID 422 422
SQ SEQUENCE 458 AA; 50872 MW; 9105479F85C6FF7 CRC64;

Query Match 53.9%; Score 1840; DB 1; Length 458;
Best Local Similarity 90.4%; Pred. No. 4,8e-110;
Matches 356; Conservative 16; Mismatches 22; Indels 0; Gaps 0;

QY 1 MNRGVPFRHLVLVQLALPAATQGNKVVYGGKGDVETLCTASOKKSIOFHKNKSNQIK 60
DB 1 MNRGIFPRHLVLVQLALPAATQGNKVVYGGKGDVETLCTASOKKSIOFHKNKSNQIK 60
QY 1LGNQGSFLTKGSPSKINDRADSRSLMDQGNFPLIINKLIEDSDTYICEVEDQKEVOL 120
DB 61 ILGIQGFLLTKGSPSKINDRADSRSLMDQGSFMIINKLIEDSDTYICEVENKKEVEL 120
QY 121 LVFGLTANSDTHLLQGSQSLTLTLSPGSSPSVQCRSPGKNIOGKTLVSQLELQDSG 180
DB 121 LVFGLTANSDTHLLQGSQSLTLTLSPGSSPSVQCRSPGKNIOGKTLVSQLELQDSG 180
QY 181 TWTCTVLQNKQKVEFKIDIVLAFQKASSIVYKKEGQVEFSPPLAFTVEKLTGSGELMW 240
DB 181 TWTCTVSQDKQKVEFKIDIVLAFQKASSIVYKKEGQVEFSPPLAFTVEKLTGSGELMW 240
QY 241 QAEARASSKSWITFDLKNKEVSVKRYTQDPKLGKGLPLHLTLPALPOYAGSGNLTLLA 300
DB 241 QAEARASSKSWITFDLKNKEVSVKRYTQDPKLGKGLPLHLTLPALPOYAGSGNLTLLA 300
QY 301 LEAKTGKLEQEVNLVVMRATQLOKNTLCEVWGSPSPKMLSLKLENKAVSKREKPVVW 360
DB 301 LEAKTGKLEQEVNLVVMRATQLOKNTLCEVWGSPSPKMLSLKLENKAVSKREKPVVW 360
QY 361 LNPEAGMWOCCLSDSGOVLLESNIKVLPTWSTPV 394
DB 361 LNPEAGMWOCCLSDSGOVLLESNIKVLPTWSTPV 394

```

```

RESULT 6
CD4_MACNE
ID CD4_MACNE STANDARD; PRT; 458 AA.
AC O08340; P79196;
DT 01-FEB-1995 (Rel. 31, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
  T4/Leu-3).
GN CD4.
OS Macaca nemestrina (pig-tailed macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.
OX NCBI_TaxID=9545;
RN [1]
RP SEQUENCE FROM N.A.
RA Hashimoto O., Tatsu M.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 28-424 FROM N.A.
RC TISSUE=Blood;
RX MEDLINE=93049640; PubMed=1425921;
RA Fomsgard A., Hirsch V.M., Johnson P.R.;
RT "Cloning and sequences of primate CD4 molecules: diversity of the
  cellular receptor for simian immunodeficiency virus/human
  immunodeficiency virus."
RL Eur. J. Immunol. 22:2973-2981(1992).
CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
  receptor interaction. May regulate T-cell activation.
CC -1- SUBUNIT: Associates with p56-lck (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.ebi.ac.uk/announcements
  or send an email to license@ebi.ac.uk).
```

```

FT CARBOHYD 42 42 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 296 296 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT CARBOHYD 325 325 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT DISULFID 41 109 BY SIMILARITY.
FT DISULFID 155 184 BY SIMILARITY.
FT DISULFID 328 370 BY SIMILARITY.
FT LIPID 419 419 S-palmitoyl cysteine (By similarity).
FT LIPID 422 422 S-palmitoyl cysteine (By similarity).
FT CONFLICT 57 57 D -> N (IN REF. 2).
FT CONFLICT 91 91 C -> H (IN REF. 2).
FT CONFLICT 105 105 N -> D (IN REF. 2).
FT CONFLICT 113 113 N -> D (IN REF. 2).
FT CONFLICT 302 302 D -> E (IN REF. 2).
FT CONFLICT 349 349 T -> A (IN REF. 2).
SQ SEQUENCE 458 AA; 50905 MW; 751A9BA2C8B3EE16 CRC64;

Query Match 53.7%; Score 1835; DB 1; Length 458;
Best Local Similarity 89.8%; Pred. No. 1e-109;
Matches 354; Conservative 19; Mismatches 21; Indels 0; Gaps 0;

QY 1 MNRGVPFRHLLVQLALLPAATQGNKVYIGKGGDYVELTCTASOKKSIQPHKNSQIK 60
DB 1 MNRGIPFRHLLVQLALLPAVTVQKKGVYIGKGGDYVELTCNASKKNTQPHKNSQIK 60
QY 61 ILGQSGFLTKGSPKLNDRASRSLMDQGNFPIIKNLKIEBSPDYICEVDQKEEVL 120
DB 61 ILGQSGFLTKGSPKLNDRASRSLMDQGNFPIIKNLKIEBSPDYICEVDQKEEVL 120
QY 121 LVFGLTANSDTHLLQGGSLTLLTSPGSSPVQCRSPKNTIQGKTLTSSVSOLEQDSG 180
DB 121 LVFGLTANSDTHLLQGGSLTLLTSPGSSPVQCRSPKNTIQGKTLTSSVSOLEQDSG 180
QY 121 LVFGLTANSDTHLLQGGSLTLLTSPGSSPVQCRSPKNTIQGKTLTSSVSOLEQDSG 180
DB 121 LVFGLTANSDTHLLQGGSLTLLTSPGSSPVQCRSPKNTIQGKTLTSSVSOLEQDSG 180
QY 181 TWICTVQONQKVEFKDIYVLAFOKASSIVYKKEGQVEFSPPLATVEKLTGSGELW 240
DB 181 TWICTVQONQKVEFKDIYVLAFOKASSIVYKKEGQVEFSPPLATVEKLTGSGELW 240
QY 181 TWICTVQONQKVEFKDIYVLAFOKASSIVYKKEGQVEFSPPLATVEKLTGSGELW 240
DB 181 TWICTVQONQKVEFKDIYVLAFOKASSIVYKKEGQVEFSPPLATVEKLTGSGELW 240
QY 241 QABRASSKSMITFDLKNKESVYKRVYQDPKLMGKKPLPHLTLPOLPOYAGSGNLTIA 300
DB 241 QABRASSKSMITFDLKNKESVYKRVYQDPKLMGKKPLPHLTLPOLPOYAGSGNLTIA 300
QY 301 LEAKTGLHGEVNVLVVVRATQLOKNTLCEVWGPTSPKMLSLKLENKAVYSKREKPVW 360
DB 301 LEAKTGLHGEVNVLVVVRATQLOKNTLCEVWGPTSPKMLSLKLENKAVYSKREKPVW 360
QY 361 LNPBAGMOCCLSDSGQVLLSNIKVLPTSTPV 394
DB 361 LNPBAGMOCCLSDSGQVLLSNIKVLPTSTPV 394

RESULT 7
CD4_CERAE
ID CD4_CERAE STANDARD; PRT; 458 AA.
AC O08338; O02805; O7593; Q28217;
DT 01-FEB-1995 (Rel. 31, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
  T4/Leu-3).
GN CD4.
OS Cercopithecus aethiops (Green monkey) (Grivet).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Cercopithecus.
OX NCBI_TaxID=9534;
RN [1]
RP SEQUENCE FROM N.A.
RA Hashimoto O., Tatsu M.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 28-424 FROM N.A.
RC TISSUE=Blood;
RX MEDLINE=93049640; PubMed=1425921;
```

RA Fomsgaard A., Hirsch V.M., Johnson P.R.:
 RT "Cloning and sequences of primate CD4 molecules: diversity of the
 RT cellular receptor for simian immunodeficiency virus/human
 RT immunodeficiency virus.";
 RL Eur. J. Immunol. 22:2973-2981(1992).
 RP [3]
 RP SEQUENCE OF 28-424 FROM N.A.
 RX TISSUE=Peripheral blood.
 RA MEDLINE=98017879; PubMed=9379478;
 RA Fomsgaard A., Mueller-Trutwin M.C., Diop O., Hansen J., Mathiot C.,
 RA Corbett S., Barte-Sinoussi F., Allan J.S.;
 RT "Relation between phylogeny of African green monkey CD4 genes and
 RT their respective simian immunodeficiency virus genes.";
 RL J. Med. Primatol. 26:120-128(1997).
 RN [4]
 RN SEQUENCE OF 107-192 FROM N.A.
 RA MEDLINE=98320644; PubMed=9656488;
 RA Harris E.E., Disotell T.R.;
 RT "Nuclear gene trees and the phylogenetic relationships of the
 RT mangabeys (primates: Papionini).";
 RL Mol. Biol. Evol. 15:892-900(1998).
 CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
 CC receptor interaction. May regulate T-cell activation.
 CC -1- SUBUNIT: Associates with p56-lck (By similarity).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>
 CC or send an email to license@sb-sib.ch).
 CC -----
 DR EMBL: D86589; BA13132.1; -;
 DR EMBL: X73322; CA51748.1; -;
 DR EMBL: AF001226; AAB60873.1; -;
 DR EMBL: AF001228; AAB60875.1; -;
 DR EMBL: AF057380; AAC25124.1; -;
 DR HSSP: P01730; 1W1Q.
 DR GO: GO:0042101; C:T-cell receptor complex; ISS.
 DR GO: GO:0015026; F:coreceptor activity; ISS.
 DR GO: GO:0042289; F:MHC class II protein binding; ISS.
 DR GO: GO:0006955; P:immune response; ISS.
 DR GO: GO:0045086; P:positive regulation of interleukin-2 biosyn. . .; ISS.
 DR GO: GO:0030217; P:T-cell differentiation; ISS.
 DR GO: GO:0045058; P:T-cell selection; ISS.
 DR GO: GO:0007169; P:transmembrane receptor protein tyrosine kin. . .; ISS.
 DR InterPro: IPR000973; CD4_TcRb.
 DR InterPro: IPR007110; Ig_Like.
 DR InterPro: IPR003596; Ig_v.
 DR Pfam: PF00047; Ig; 2.
 DR PRINTS: PRO0692; CD4TCNTGEN.
 DR SMART: SM00406; IGV; 1.
 DR PROSITE: PS50835; IG_Like; 1.
 KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
 KW Immune response; Repeat; Signal; Lipoprotein; Palmitate.
 FT SIGNAL 1 25
 FT CHAIN 26 458
 FT DOMAIN 26 396
 FT TRANSMEM 397 418
 FT DOMAIN 419 458
 FT DOMAIN 126 125
 FT DOMAIN 203 317
 FT DOMAIN 318 374
 FT DOMAIN 374 421
 FT CARBOHYD 42 42
 FT CARBOHYD 281 281
 FT CARBOHYD 296 296
 FT CARBOHYD 325 325
 FT DISULFID 41 109
 FT BY SIMILARITY.

FT DISULFID 155 184 BY SIMILARITY.
 FT DISULFID 328 370 BY SIMILARITY.
 FT LIPID 419 439 S-palmitoyl cysteine (By similarity).
 FT LIPID 422 422 S-palmitoyl cysteine (By similarity).
 FT CONFLICT 46 46 K -> N (IN REF. 2 AND 3; AAB60875).
 FT CONFLICT 59 59 K -> T (IN REF. 3; AAB60873).
 FT CONFLICT 115 115 K -> E (IN REF. 1).
 FT CONFLICT 165 165 G -> V (IN REF. 3; AAB60873 AND 4).
 FT CONFLICT 200 200 M -> V (IN REF. 2 AND 3).
 FT CONFLICT 227 227 F -> L (IN REF. 3; AAB60873).
 FT CONFLICT 271 271 K -> E (IN REF. 2; AAB60873).
 FT CONFLICT 281 281 N -> H (IN REF. 3; AAB60873).
 SQ SEQUENCE 458 AA; 51158 MW; FCS23D2EDD1F72E7 CRC64;
 Query Match 53.3%; Score 1818; DB 1; Length 458;
 Best Local Similarity 89.3%; Pred. No. 1.2e-108; Indels 0; Gaps 0;
 Matches 352; Conservative 17; Mismatches 25;
 QY 1 MNRGVFRLHLVQLALPAATQGNKVLGKKGDTVELTCTASQKSIQFHKNSNQIK 60
 DB 1 MNMGIFRHLVLQALPAVATQKVLGKKGDTVELTCTASQKSIQFHKNSNQIK 60
 QY 61 IIGNGSFLTKGPSKLNDRADSRISIMDGNFPLIKLKIEDSDTYICEVEDQKEEYOL 120
 DB 61 IIGKGSFLTKGSSKLRDRIDSRKSLMDQCFSMIKNLKIEDSEFYICEVENKKEEVEI 120
 QY 121 LVFGLTANSDTHLQGSITLTLSPGSSPSVQCRSPGKXNIGGKTLVSQLEIDSG 180
 DB 121 LVFGLTANSDTHLQGSITLTLSPGSSPSVQCRSPGKXNIGGKTLVSQLEIDSG 180
 QY 181 TWTCYTLQNKQKVEFDIVLAFQKASSIVKKEGEQVEFPLAFYTEKLTGSGELMW 240
 DB 181 TWTCYTLQNKQKVEFDIVLAFQKASSIVKKEGEQVEFPLAFYTEKLTGSGELMW 240
 QY 241 QERASSSSKSWITFDLKNKESVVKRYTQDPKLQMGKPLPLHLTLPALPQYAGSGNLTLA 300
 DB 241 QERASSSSKSWITFDLKNKESVVKRYTQDPKLQMGKPLPLHLTLPALPQYAGSGNLTLA 300
 QY 301 LEAKTKLHOENLVLMRATLOKNTCEVMPPTSPKMLSLKLEKKAQVSRERQVWV 360
 DB 301 LEAKTKLHOENLVLMRATLOKNTCEVMPPTSPKMLSLKLEKKAQVSRERQVWV 360
 QY 361 LNPEAGMOCLSDSGVLLLESINIKVLPMTSTPV 394
 DB 361 LNPEAGMOCLSDSGVLLLESINIKVLPMTSTPV 394
 RESULT 8
 ID CD4_ERYPA STANDARD; PRT; 397 AA.
 AC 008339;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/Leu-3)
 DE (Fragment).
 GN CD4.
 OS Erythrocybus patas (Red guenon) (Husar).
 OC Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecoidea; Erythrocybus.
 NCBI_Taxid=9538;
 RN [1]
 RP SEQUENCE FROM N.A.
 RP TISSUE=Blood;
 RX MEDLINE=93049640; PubMed=1425921;
 RX Fomsgaard A., Hirsch V.M., Johnson P.R.;
 RT "Cloning and sequences of primate CD4 molecules: diversity of the
 RT cellular receptor for simian immunodeficiency virus/human
 RT immunodeficiency virus.";
 RL Eur. J. Immunol. 22:2973-2981(1992).
 CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
 CC receptor interaction. May regulate T-cell activation.

```

CC -1- SUBUNIT: Associates with p56-lck (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: X73324; CAAS1750.1; -.
DR HSSP: P01730; IMIO.
DR GO: GO:0042101; C:T-cell receptor complex; ISS.
DR GO: GO:0015026; F:coreceptor activity; ISS.
DR GO: GO:0042289; F:MHC class II protein binding; ISS.
DR GO: GO:0006955; P:immune response; ISS.
DR GO: GO:0045086; P:positive regulation of interleukin-2 biosyn. . . ; ISS.
DR GO: GO:0030217; P:T-cell differentiation; ISS.
DR GO: GO:0045058; P:T-cell selection; ISS.
DR GO: GO:0007169; P:transmembrane receptor protein tyrosine kin. . . ; ISS.
DR InterPro: IPR000973; CD4_TcAg.
DR InterPro: IPR007110; Ig_Like.
DR Pfam: PF00047; Ig_2.
DR PRINTS: PR00692; CD4TCANTIGEN.
DR SMART: SM00406; IGV; 1.
DR PROSITE: PS50835; IG_LIKE; 1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Immune response; Repeat; Lipoprotein; Palmitate.
FT NON_TER 1 1
FT DOMAIN 370 391 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 370 391 POTENTIAL.
FT CYTOPLASMIC (POTENTIAL).
FT DOMAIN 392 >397 IG-LIKE V-TYPE.
FT DOMAIN 392 >397 IG-LIKE V-TYPE.
FT DOMAIN 392 >397 IG-LIKE C2-TYPE 1.
FT DOMAIN 392 >397 IG-LIKE C2-TYPE 2.
FT DOMAIN 392 >397 IG-LIKE C2-TYPE 3.
FT CARBOHYD 269 269 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT CARBOHYD 298 298 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT DISULFID 14 82 BY SIMILARITY.
FT DISULFID 128 157 BY SIMILARITY.
FT DISULFID 301 343 BY SIMILARITY.
FT LIPID 392 392 S-palmitoyl cysteine (By similarity).
FT LIPID 395 395 S-palmitoyl cysteine (By similarity).
FT NON_TER 397 397
SQ SEQUENCE 397 AA; 44081 MW; 67887397A6B7EA4F CRC64;
-----
Query Match 50.3%; Score 1717; DB 1; Length 397;
Beet Local Similarity 89.9%; Pred. No. 2.7e-102;
Matches 330; Conservative 15; Mismatches 22; Indels 0; Gaps 0;

```

```

QY 328 CEVAGPTSPKLMISLKENKEAKVSKREKPVWVLPAGWMOCLSDSGVLTESNIVL 387
DB 301 CEVAGPTSPKLTSLKENKEATISKOAKAVWVLPBEGWMOCLSDSGVLTESNIVL 360
QY 388 PTWSTPV 394
DB 361 PTWPTPV 367
-----
RESULT 9
CD4_CERTO STANDARD; PRT; 397 AA.
ID CD4_CERTO
AC 008336;
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/Leu-3)
DE (Fragment).
GN CD4.
OS Cercopithecus torquatus atys (Red-crowned mangabey) (Sooty mangabey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Cercopithecus.
OC NCBI_TaxID=9531;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Blood;
RX MEDLINE=93049640; PubMed=1425921;
RA Fomsgaard A., Hirsch V.M., Johnson P.R.;
RT "Cloning and sequences of primate CD4 molecules: diversity of the
RT cellular receptor for simian immunodeficiency virus/human
RT immunodeficiency virus";
RL Eur. J. Immunol. 22:2973-2981(1992).
RL
CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
CC receptor interaction. May regulate T-cell activation.
CC -1- SUBUNIT: Associates with p56-lck (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: X73328; CAAS1754.1; -.
DR EMBL: X73327; CAAS1753.1; -.
DR HSSP: P01730; IMIO.
DR GO: GO:0042101; C:T-cell receptor complex; ISS.
DR GO: GO:0015026; F:coreceptor activity; ISS.
DR GO: GO:0042289; F:MHC class II protein binding; ISS.
DR GO: GO:0006955; P:immune response; ISS.
DR GO: GO:0045086; P:positive regulation of interleukin-2 biosyn. . . ; ISS.
DR GO: GO:0030217; P:T-cell differentiation; ISS.
DR GO: GO:0045058; P:T-cell selection; ISS.
DR GO: GO:0007169; P:transmembrane receptor protein tyrosine kin. . . ; ISS.
DR InterPro: IPR000973; CD4_TcAg.
DR InterPro: IPR007110; Ig_Like.
DR InterPro: IPR003596; Ig_V.
DR Pfam: PF00047; Ig_2.
DR PRINTS: PR00692; CD4TCANTIGEN.
DR SMART: SM00406; IGV; 1.
DR PROSITE: PS50835; IG_LIKE; 1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Immune response; Repeat; Lipoprotein; Palmitate.
FT NON_TER 1 1
FT DOMAIN 370 391 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 370 391 POTENTIAL.
FT DOMAIN 392 >397 CYTOPLASMIC (POTENTIAL).

```

```

FT DOMAIN <1 98 IG-LIKE V-TYPE.
FT DOMAIN 99 176 IG-LIKE C2-TYPE 1.
FT DOMAIN 177 290 IG-LIKE C2-TYPE 2.
FT DOMAIN 291 347 IG-LIKE C2-TYPE 3.
FT CARBOHYD 15 15 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 269 269 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 298 298 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DISULFID 14 82 BY SIMILARITY.
FT DISULFID 128 157 BY SIMILARITY.
FT DISULFID 301 343 BY SIMILARITY.
FT LIPID 392 392 S-palmitoyl cysteine (By similarity).
FT LIPID 395 395 S-palmitoyl cysteine (By similarity).
FT VARIANT 20 20 MISSING.
FT VARIANT 43 43 T -> I.
FT VARIANT 86 86 N -> D.
FT VARIANT 96 96 F -> L.
FT VARIANT 173 173 V -> M.
FT VARIANT 316 316 R -> K.
FT NON TER 397 397
SQ SEQUENCE 397 AA; 43926 MW; 8660B636D2B38A7 CRC64;

Query Match 50.0%; Score 1707; DB 1; Length 397;
Best Local Similarity 89.6%; Pred. No. 1.2e-101;
Matches 329; Conservative 18; Mismatches 20; Indels 0; Gaps 0;

QY 28 VVLGKGDVLTCTASQKSIQFHMKNSQIKILNQSGFLTKGSKLNDRADSRSLM 87
DB 1 VVLGKGDVLTCTASQKSIQFHMKNSQIKILNQSGFLTKGSKLNDRADSRSLM 60
QY 88 DQGNFLLIKNKLESDTYICEVEDKEEVOLLVFGLTANSPHLLQGSLLTLESPP 147
DB 61 DQGNFLLIKNKLESDTYICEVEDKEEVOLLVFGLTANSPHLLQGSLLTLESPP 120
QY 148 GSSPSVQCRSPRGKNIQSGKTLVSQLELQDSGTWTCTVLQNGKVEFKIDIVLAFQKA 207
DB 121 GSSPSVQCRSPRGKNIQSGKTLVSQLELQDSGTWTCTVLQNGKVEFKIDIVLAFQKA 180
QY 208 SSIYKKEGQVSEFPLATVTEKLTSSGELMWAQEAASSKSWITFDLKNKEVSVKRYT 267
DB 181 SSIYKKEGQVSEFPLATVTEKLTSSGELMWAQEAASSKSWITFDLKNKEVSVKRYT 240
QY 268 QDPPLQMGKPLHLTLPLALPOYAGSGNLTALAEATGKLGHOEVLVWRATQLOKNT 327
DB 241 QDPPLQMGKPLHLTLPLALPOYAGSGNLTALAEATGKLGHOEVLVWRATQLOKNT 300
QY 328 CEVWGPTSPKLMSTLKNKEAKVSKREKPVWVLPNPAQMOCLSDSGVLLSENIKVL 387
DB 301 CEVWGPTSPKLMSTLKNKEAKVSKREKPVWVLPNPAQMOCLSDSGVLLSENIKVL 360
QY 388 PTWSTPV 394
DB 361 PTWSTPV 367

RESULT 10
CD4-SAISC STANDARD; PRT; 457 AA.
AC Q29037;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
DE T4/Leu-3).
GN CD4.
OS Saimiri sciureus (Common squirrel monkey).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Saimiri.
OX NCBI_TaxID=95211;
RN [1]
RP SEQUENCE FROM N.A.
RA Tacsuni M., Hashimoto O.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
-!- FUNCTION: Accessory protein for MHC class-II antigen/T-cell

```

```

CC receptor interaction. May regulate T-cell activation.
CC -!- SUMMARY: Associates with p56-lck (By similarity).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D86588; BAA13131.1; -.
CC HSSP: P01730; IMR.
CC DR GO: GO:0042101; C: T-cell receptor complex; ISS.
CC DR GO: GO:0015026; P: coreceptor activity; ISS.
CC DR GO: GO:0042289; F: MHC class II protein binding; ISS.
CC DR GO: GO:0006955; P: immune response; ISS.
CC DR GO: GO:0045086; P: positive regulation of interleukin-2 biosyn. .; ISS.
CC DR GO: GO:0030217; P: T-cell differentiation; ISS.
CC DR GO: GO:0045058; P: T-cell selection; ISS.
CC DR GO: GO:0071659; P: transmembrane receptor protein tyrosine kin. .; ISS.
CC DR InterPro: IPR000973; CD4-TCRG.
CC DR InterPro: IPR007110; Ig-Like.
CC DR InterPro: IPR003596; Ig_V.
CC DR Pfam: PF00047; Ig_2.
CC DR PRINTS: PR00692; CD4TCANTIGEN.
CC DR SMART: SM00406; IGV; 1.
CC DR PROSITE: PS00835; IG_LIKE; 1.
CC KM Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
CC KM Immune response; Repeat; Signal; Lipoprotein; Palmitate.
CC FT CHAIN 1 25
CC FT DOMAIN 26 457
CC FT TRANSMEM 396 417
CC FT DOMAIN 418 457
CC FT DOMAIN 426 457
CC FT DOMAIN 126 202
CC FT DOMAIN 203 316
CC FT DOMAIN 317 373
CC FT CARBOHYD 254 254 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 295 295 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 324 324 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT DISULFID 41 109 BY SIMILARITY.
CC FT DISULFID 155 184 BY SIMILARITY.
CC FT DISULFID 327 369 BY SIMILARITY.
CC FT LIPID 418 418 S-palmitoyl cysteine (By similarity).
CC FT LIPID 421 421 S-palmitoyl cysteine (By similarity).
CC SQ SEQUENCE 457 AA; 50871 MW; 57EBD6344005A015 CRC64;

Query Match 46.1%; Score 1572.5; DB 1; Length 457;
Best Local Similarity 78.2%; Pred. No. 5.2e-93;
Matches 308; Conservative 31; Mismatches 54; Indels 1; Gaps 1;

QY 1 MNRGVFPHLLVLTQALLPAATQGNKVLGKGGDVELTCTASQKSIQFHMKNSQIK 60
DB 1 MNRGVFPHLLVLTQALLPAATQGNKVLGKGGDVELTCTASQKSIQFHMKNSQIK 60
QY 61 ILGNQSPFLTKGSKLNDRADSRSLMDQGNFLLIKNKLESDTYICEVEDKEEVOLL 120
DB 61 ILGNQSPFLTKGSKLNDRADSRSLMDQGNFLLIKNKLESDTYICEVEDKEEVOLL 120
QY 121 LVFGLTANSPHLLQGSLLTLESPPGSSPVQCRSPRGKNIQSGKTLVSQLELQDSG 180
DB 121 QVWGTLTANSPHLLQGSLLTLESPPGSSPVQCRSPRGKNIQSGKTLVSQLELQDSG 180
QY 181 TWCTCTVLQNGKVEFKIDIVLAFQKASSIVYKKEGQVSEFPLATVTEKLTSSGEL 240
DB 181 TWCTCTVLQNGKVEFKIDIVLAFQKASSIVYKKEGQVSEFPLATVTEKLTSSGEL 239
QY 241 QABRASSKSWITFDLKNKEVSVKRYTQDPKLMGKPLHLTLPLALPOYAGSGNLT 300

```

```

DB 240 QAEKSSSSKSWTFNLTQOEYVVKLVTDPKLRMGKKLPMLHTLAQALPOYAGSGNFTLA 299
QY 301 LEAKTGKHOEYNLVVMBATOLQKRLTCEVWGPTSPKMLSLKLENKAKYSKREKPVW 360
DB 300 LKGTGKHQOEYNLVVMBATOLQKRLTCEVWGPTSPKMLSLKLENKAKYSKREKAVW 359
QY 361 LNPEAGWMOCLLSDSGQVLLSEINIKVLPWTSTPV 394
DB 360 LNPEAGWMOCLLSDSGQVLLSEKFEALPTRSPPV 393

RESULT 11
GCL HUMAN STANDARD; PRT; 330 AA.
AC P01857;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ig gamma-1 chain C region.
GN IGHG1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OC NCBI_TaxID=9606;
RX CUNNINGHAM B.A., RUTISHAUSER U., GALL W.E., GOTTLIEB P.D.,
RX WAXDAL M.J., EDELMAN G.M.;
RT "The covalent structure of a human gamma G-immunoglobulin. VII. Amino
RL Biochemistry 9:3161-3170 (1970).
[3]
RX MEDLINE=136-329 (EU).
RX MEDLINE=71064025; PubMed=5530842;
RA Rutishauser U., Cunningham B.A., Bennett C., Konigsberg W.H.,
RA Edelman G.M.;
RT "The covalent structure of a human gamma G-immunoglobulin. 8. Amino
RL Biochemistry 9:3171-3181 (1970).
[4]
RX SEQUENCE (MYELOMA PROTEIN NIE).
RX MEDLINE=71070269; PubMed=826475;
RA Ponstingl H., Hilschmann N.;
RT "The rule of antibody structure. The primary structure of a
RT monoclonal IgG1 immunoglobulin (myeloma protein Nie). III. The
RT tryptic peptides of the H-chain, alignment of the tryptic
RL Hoppe-Seyler's Z. Physiol. Chem. 357:1571-1604 (1976).
[5]
RX SEQUENCE (MYELOMA PROTEIN KOL), AND DISULFIDE BONDS.
RX MEDLINE=83289131; PubMed=6884994;
RA Schmidt W.E., Jung H.-D., Palm W., Hilschmann N.;
RT "Three-dimensional structure determination of antibodies. Primary
RT structure of crystallized monoclonal immunoglobulin IgG1 KOL, I.",
RL Hoppe-Seyler's Z. Physiol. Chem. 364:713-747 (1983).
[6]
RX DISULFIDE BONDS.
RX MEDLINE=71064027; PubMed=4923144;
RA Gall W.E., Edelman G.M.;
RT "The covalent structure of a human gamma G-immunoglobulin. X.
RL Interchain disulfide bonds.",
RL Biochemistry 9:3188-3196 (1970).
[7]
RX DISULFIDE BONDS.
RX MEDLINE=71070267; PubMed=1002129;

```

```

RA Dreker L., Schwarz J., Reichel W., Hilschmann N.;
RT "Rule of antibody structure. The primary structure of a monoclonal
RT IgG1 immunoglobulin (myeloma protein Nie). I: Purification and
RT characterization of the protein, the L- and H-chains, the
RL cytochrome brome cleavage products, and the disulfide bridges.",
RL Hoppe-Seyler's Z. Physiol. Chem. 357:1515-1540 (1976).
[8]
RX X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS).
RX MEDLINE=81208100; PubMed=7236608;
RA Drenth J.;
RT "Crystallographic refinement and atomic models of a human Fc fragment
RT and its complex with fragment B of protein A from Staphylococcus
RL aureus at 2.9- and 2.8-A resolution.",
RL Biochemistry 20:2361-2370 (1981).
CC -I- MISCELLANEOUS: Nie has the GIM(17) allotypic marker, 97-K, and the
CC GIM(1) markers, 239-D and 241-L. KOL and EU sequences have the
CC GIM(3) marker and the GIM (non-1) markers.
CC -I- MISCELLANEOUS: Nie also differs in the amidation states of
CC 35, 116, 198, 269 and 272.
CC -I- MISCELLANEOUS: EU also differs in the amidation states of residues
CC 155, 166, 177, 195, 198, 269, and 272 and in the order of residues
CC 268-272.
CC -I- MISCELLANEOUS: KOL also differs in the amidation states of
CC residues 198, 267 and 272.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
-----
DR EMBL, J00228; AAC82527.1; ALT_INIT.
DR PIR, A93433; GHNU.
DR PDB, 1FC1; 15-JUL-92.
DR PDB, 1FC2; 15-JUL-92.
DR PDB, 1AJ7; 12-NOV-97.
DR PDB, 1D57; 09-FEB-00.
DR PDB, 1D51; 09-FEB-00.
DR PDB, 1D6V; 04-OCT-00.
DR PDB, 1DN2; 17-MAY-00.
DR PDB, 1E4K; 06-JUN-01.
DR PDB, 1FCC; 20-JUL-95.
DR PDB, 1H2H; 12-JUN-02.
DR PDB, 1I7Z; 08-AUG-01.
DR PDB, 1I1S; 16-MAY-01.
DR PDB, 1I1X; 16-MAY-01.
DR PDB, 1L6X; 10-APR-02.
DR PDB, 2RCS; 12-NOV-97.
DR Genew; HGNC:5525; IGHG1.
DR MIM, 147100; -.
DR GO; GO:0005624; C:membrane fraction; NAS.
DR GO; GO:0003823; F:antigen binding; TAS.
DR GO; GO:0006955; P:immune response; NAS.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003597; Ig_c1.
DR Pfam; PF00047; Ig_3.
DR SMART; SM00407; IGH1; 2.
DR PROSITE; PS00835; IG_LIKE; 3.
DR PROSITE; PS00290; IG_MHC; 2.
KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
KW 3D-structure.
FT NON TER 1 1
FT DOMAIN 1 98 CH1.
FT DOMAIN 99 110 HINGE.
FT DOMAIN 111 223 CH2.
FT DOMAIN 224 330 CH3.
FT DISULFID 27 83
FT DISULFID 103 103 INTERCHAIN (WITH LIGHT CHAIN).
FT DISULFID 109 109 INTERCHAIN (WITH HEAVY CHAIN).
FT DISULFID 112 112 INTERCHAIN (WITH HEAVY CHAIN).

```

```
FT DISULFID 144 204
FT DISULFID 250 308
FT CARBOHYD 180 180
FT MOD RES 330 330
FT VARIANT 97 97
FT VARIANT 239 239
FT VARIANT 241 241
FT STRAND 122 126
FT HELIX 130 134
FT TURN 136 137
FT STRAND 141 147
FT STRAND 157 162
FT TURN 163 164
FT STRAND 165 166
FT TURN 168 171
FT STRAND 176 179
FT TURN 180 181
FT STRAND 182 190
FT HELIX 193 197
FT TURN 198 199
FT STRAND 202 207
FT TURN 209 210
FT STRAND 215 219
FT STRAND 227 227
FT STRAND 230 234
FT HELIX 238 242
FT STRAND 245 256
FT STRAND 260 265
FT STRAND 270 276
FT STRAND 274 276
FT TURN 280 281
FT STRAND 283 284
FT STRAND 287 296
FT HELIX 297 301
FT TURN 302 303
FT STRAND 305 312
FT TURN 313 314
FT TURN 316 317
FT STRAND 320 325
SQ SEQUENCE 330 AA; 36106 MM; 3770EHL06C2FA33D CRC64;

Query Match 35.5%; Score 1211.5; DB 1; Length 330;
Best Local Similarity 97.8%; Pred. No. 3.3e-70;
Matches 226; Conservative 0; Mismatches 0; Indels 5; Gaps 1;

QY 400 EPKSCDKTHTC-----PELLGSPVFLPPKPKDTLMISRTPEVTCVVVDSHEDPEVKF 454
DB 99 EPKSCDKTHTCPCPAPPELLGGPSVFLFPPPKDTLMISRTPEVTCVVVDSHEDPEVKF 158
QY 455 NMVYDGEVHNAAKTPREBOYNSTYRVSVLTVLHODMLNGKKEKCKVSNKALPAPIEKT 514
DB 159 NMVYDGEVHNAAKTPREBOYNSTYRVSVLTVLHODMLNGKKEKCKVSNKALPAPIEKT 218
QY 515 ISRAKGPREPQVYTLPPSDELTKNOVSLTCLVKGFPSPDIAVEMESNQPENNYKTP 574
DB 219 ISRAKGPREPQVYTLPPSDELTKNOVSLTCLVKGFPSPDIAVEMESNQPENNYKTP 278
QY 575 PVLDSDSFLLSKLTVDKSRMOGNVFSQVMEALAHNYTKSLSPG 625
DB 279 PVLDSDSFLLSKLTVDKSRMOGNVFSQVMEALAHNYTKSLSPG 329

RESULT 12
CD4_RABIT STANDARD; PRT; 459 AA.
AC P46630;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
```

```
DE 74/Leu-3).
GN CD4.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_Taxid:9986;
OC [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92390370; PubMed=1518821;
RA Hogue B.F., Sawaadikosol S., Brown T.J., Lee K., Recker D.P.,
RA Kande T.J.;
RT "CD4 and its role in infection of rabbit cell lines by human
RT immunodeficiency virus type 1."
RL Proc. Natl. Acad. Sci. U.S.A. 89:7963-7967(1992).
CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
CC receptor interaction. May regulate T-cell activation.
CC -1- SUBUNIT: Associates with p56-lck (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M92840; AAA1198.1; -.
CC PIR; A46254; A46254.
CC HSSP; P01730; 1MBR.
DR GO; GO:0042101; C: T-cell receptor complex; ISS.
DR GO; GO:0015026; F: coreceptor activity; ISS.
DR GO; GO:0042289; F: MHC class II protein binding; ISS.
DR GO; GO:0069555; P: immune response; ISS.
DR GO; GO:0045086; P: positive regulation of interleukin-2 biosyn. . .; ISS.
DR GO; GO:0030217; P: T-cell differentiation; ISS.
DR GO; GO:0045058; P: T-cell selection; ISS.
DR GO; GO:0007169; P: transmembrane receptor protein tyrosine kin. . .; ISS.
DR InterPro; IPR000973; CD4_TCMG.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PR00047; Ig_2.
DR PRINTS; PR00692; CD4TCANTIGEN.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
KM Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Immune response; Repeat; Signal; Lipoprotein; Palmitate.
FT SIGNAL 1 25
FT CHAIN 26 459
FT DOMAIN 26 396
FT TRANSMEM 397 419
FT DOMAIN 420 459
FT DOMAIN 26 129
FT DOMAIN 130 208
FT DOMAIN 209 318
FT DOMAIN 319 374
FT CARBOHYD 299 299
FT DISULFID 41 113
FT DISULFID 329 370
FT LIPID 420 420
FT LIPID 423 423
SQ SEQUENCE 459 AA; 50886 MM; B323311CB040013D CRC64;

Query Match 33.7%; Score 1150.5; DB 1; Length 459;
Best Local Similarity 57.2%; Pred. No. 3.9e-66;
Matches 241; Conservative 65; Mismatches 90; Indels 25; Gaps 6;

QY 1 MNRGVEFRHLVLTQLALLPAATQGNKVVYLGKKGDVLELTCTASQKSKSIQFHKNNSNQIK 60
DB 1 MNRRIYFGCLLVLPALLPAATWGKTVVRGKAGALVELPCQSSQKNSVFNWGHANQVK 60
```



```

FT          VARIANT      279      279      /FTId=VAR_003895.
FT          SEQUENCE      290 AA; 32331 MW; E69CBG95705B2F46 CRC64;
SQ
Query Match      33.5%; Score 1143; DB 1; Length 290;
Best Local Similarity 87.9%; Pred. No. 6,4e-66;
Matches 211; Conservative 13; Mismatches 10; Indels 6; Gaps 2;

QY          392 TVVPCP-APPEKSCDKTHTC-----PELLGSPVFLPPPKDTLMISRTPEVTCVVVDY 445
DB          50 TPEPCRCPEPKSCDTPPPPCPCPAPELLGSPVFLPPPKDTLMISRTPEVTCVVVDY 109
QY          446 SHEDPEVKFWYVDGVVHNAKTKRPREQVNSTYRVAVSVTLVHQMNLNKEVKCKVSNK 505
DB          110 SHEDPEVQFKWYVDGVVHNAKTKRPRQVNSTFRVAVSVTLVHQMNLNKEVKCKVSNK 169
QY          506 ALPAPIEKTSKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEMESNQ 565
DB          170 ALPAPIEKTSKTKGQPREPQVYTLPPSRDEMTKNQVSLTCLVKGFYPSDIAVEMESSGQ 229
QY          566 PENNYKTPPVLDSDGSFPLYSKLTVDKSRWQQGNVFCGVMHEDALHNNHYTKSLSPG 625
DB          230 PENNYKTPPVLDSDGSFPLYSKLTVDKSRWQQGNVFCGVMHEDALHNNHYTKSLSPG 289

RESULT 14
CD4 CANFA ID CD4 CANFA STANDARD; PRT; 463 AA.
AC P33705;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
   T4/Lew-3).
GN CD4.
OS Canis familiaris (Dog).
OC Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN
RP SEQUENCE OF 13-463 FROM N.A.
RC STRAIN=Beagle; TISSUE=Thymus;
RX MEDLINE=93192324; Pubmed=791632;
RA Milde K.F., Conner G.E., Minz D.H., Alejandro R.;
RT "Primary structure of the canine CD4 antigen.";
RL Biochim. Acta 1172:315-318(1993).
RN
RP SEQUENCE FROM N.A.
RC STRAIN=Beagle; TISSUE=Thymus;
RX MEDLINE=94378217; Pubmed=8091416;
RA Gorman S.D., Fretwin M.R., Cobbold S.P., Waldmann H.;
RT "Isolation and expression of cDNA encoding the canine CD4 and CD8
   alpha antigens.";
RL Tissue Antigens 43:184-188(1994).
CC -|- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
   receptor interaction. May regulate T-cell activation.
CC -|- SUBUNIT: Associates with p56-lck (by similarity).
CC -|- SUBCELLULAR LOCATION: Type I membrane protein.
CC -|- TISSUE SPECIFICITY: Expressed in macrophages and a subset of
   T lymphocytes.
CC -|- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -|- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; L06130; AAB02295.1; -

```

```

DR          EMBL; X68565; -; NOT_ANNOTATED_CDS.
DR          HSSP; P01730; 1MR.
DR          GO; GO:0042101; C1T-cell receptor complex; ISS.
DR          GO; GO:0015026; F:coreceptor activity; ISS.
DR          GO; GO:0042289; F:MHC class II protein binding; ISS.
DR          GO; GO:0006955; P:immune response; ISS.
DR          GO; GO:0045086; P:positive regulation of interleukin-2 biosyn. . .; ISS.
DR          GO; GO:0030217; P:T-cell differentiation; ISS.
DR          GO; GO:0045058; P:T-cell selection; ISS.
DR          GO; GO:0007169; P:transmembrane receptor protein tyrosine kin. . .; ISS.
DR          InterPro; IPR000973; CD4_TcRg.
DR          InterPro; IPR007110; Ig-like.
DR          InterPro; IPR003596; Ig_V.
DR          Pfam; PF00047; Ig_3.
DR          PRINTS; PR00692; CD4TCANTIGEN.
DR          SMART; SM00406; IGV; 1.
DR          PROSITE; PS50835; IG_LIKE; 1.
DR          KMW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
   KM Immune response; Repeat; Signal; Lipoprotein; Palmitate.
FT          SIGNL 1 24
FT          CHAIN 25 463
FT          DOMAIN 25 401
FT          TRANSMEM 402 423
FT          DOMAIN 424 463
FT          DOMAIN 26 124
FT          DOMAIN 125 211
FT          DOMAIN 212 321
FT          DOMAIN 322 378
FT          DISULFID 374 109
FT          LIPID 424 424
FT          LIPID 427 427
FT          CARBOHYD 123 123
FT          CARBOHYD 168 168
FT          CARBOHYD 176 176
FT          CARBOHYD 324 324
FT          CARBOHYD 329 329
FT          CARBOHYD 389 389
SO          SEQUENCE 463 AA; 51639 MW; 95805170CB4A833 CRC64;

Query Match      33.2%; Score 1135; DB 1; Length 463;
Best Local Similarity 57.8%; Pred. No. 3,8e-65;
Matches 233; Conservative 63; Mismatches 89; Indels 18; Gaps 6;

QY          1 NMRGVFRLHLVQLALLPATQGNKYVLGKGDVVELTCTASOKKSIQFHKNKNQIK 60
DB          1 MNQEAFFRHLMLQLVMLPAVTPVREVLGKAGDAVELPCQTSQKKNIHFNRDSSMVQ 60
QY          61 IICNGSFLTKGPKSKLNDRADRSRLMDQGNFPLIKNKIEDSDTYICEVEDQKEEVQL 120
DB          61 IICNGSFWTVGSSRLKHKVESKKNIMDQSPFLVKDELVDAGSGLYFCDT-DKRGVEL 119
QY          121 LVFGILTA-----NSDTHLLQGSLLTLLESPPGSSPSVQCRSPRGKNIQSGKTLV 171
DB          120 LVFNILAKWDSCSSGSSNIRLLQGOQLTLLENPGSSPSVQWKGPKNKSKEGCONLSL 179
QY          172 SOLELDSGTWTCTVQONKVEFKDIYVLAFOKXSLIVYKKEGQVRESFLATVEK 231
DB          180 SWELEDGGGTWCTTISQSKTVEFNINVLVLAFOKXSNFYAEGSQVRESFSLSPEDEN 239
QY          232 LTGSGELMWAERASSKSMITFDLKNKEVSVKRVTDPLQMGKGLPLHLTPQALPOY 291
DB          240 LV-GEIRQDAQASSSLNISTLENKLSMEKALAPLQKESLPLRFTLPVLSRY 297
QY          292 AGSGNLTLLAEATKGLHGEVNLVVRATQLOKNLTCEVWGPTSPKLMSLKLENKAVY 351
DB          298 AGSGILTLNL-AK-GTLVGEVNLVVRANSSQNLLTCEVLGPTSPBELTSLNLKEQAKV 355
QY          352 SKREKEVWLVNPEAGMWQCLSDSGVLLSINKVLPMTSTVP 394
DB          356 SKQKLVWVVDPEGTWQCLSDKDKVLLASSLNV-----SSPV 394

```

RESULT 15
ID GC2 HUMAN STANDARD; PRT: 326 AA.
AC P01859;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ig gamma-2 chain C region.
GN IGHG2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE OF 2-326 FROM N.A.
RX MEDLINE=82197621; PubMed=6604948;
RA Ellison J.W., Hood L.E.;
RT "Linkage and sequence homology of two human immunoglobulin gamma heavy chain constant region genes.";
RL Proc. Natl. Acad. Sci. U.S.A. 79:1984-1988(1982).
RN [2]
RP SEQUENCE OF 88-115 FROM N.A.
RX TISSUE=Fetal liver;
RA MEDLINE=83001943; PubMed=6811339;
RA Takahashi N., Ueda S., Obata M., Nikaido T., Nakai S., Honjo T.;
RT "Structure of human immunoglobulin gamma genes: implications for evolution of a gene family.";
RL Cell 29:671-679(1982).
RN [3]
RP SEQUENCE OF 99-177 AND 310-326 FROM N.A.
RX TISSUE=Fetal liver;
RA MEDLINE=84235992; PubMed=6329676;
RA Krawinkel U., Rebhitzer T.H.;
RT "Comparison of the hinge-coding segments in human immunoglobulin gamma heavy chain genes and the linkage of the gamma 2 and gamma 4 subclass genes.";
RL EMBO J. 1:403-407(1982).
RN [4]
RP SEQUENCE OF 1-325 (MYELOMA PROTEIN TIL).
RX MEDLINE=81007873; PubMed=6774012;
RA Wang A.-C., Tung E., Fudenberg H.H.;
RT "The primary structure of a human IgG2 heavy chain: genetic, evolutionary, and functional implications.";
RL J. Immunol. 125:1048-1054(1980).
RN [5]
RP SEQUENCE OF 1-85 AND 132-325 (MYELOMA PROTEIN ZIE).
RX MEDLINE=80001357; PubMed=113060;
RA Conneil G.E., Parr D.M., Hofmann T.;
RT "The amino acid sequences of the three heavy chain constant region domains of a human IgG2 myeloma protein.";
RL Can. J. Biochem. 57:758-767(1979).
RN [6]
RP SEQUENCE OF 238-275 (ZIE).
RX MEDLINE=8011419; PubMed=118920;
RA Hofmann T., Parr D.M.;
RT "A note of the amino acid sequence of residues 381-391 of human immunoglobulin gamma chains.";
RL Mol. Immunol. 16:923-925(1979).
RN [7]
RP REVISIONS TO 25; 59; 60 AND 264-268 (ZIE).
RA Hofmann T., Parr D.M.;
RL Submitted (MAR-1980) to the PIR data bank.
RN [8]
RP SEQUENCE OF 1-121 (DOT).
RX MEDLINE=95255298; PubMed=7737190;
RA Scoppini M., Bellotti V., Negri A., Merlini G., Garver F., Ferri G.;
RT "Characterization of the two unique human anti-flavin monoclonal immunoglobulins.";
RL Eur. J. Biochem. 228:886-893(1995).
RN [9]
RP DISULFIDE BONDS.
RX MEDLINE=72033500; PubMed=4940472;
RA Milstein C., Frangione B.;

RT "Disulphide bridges of the heavy chain of human immunoglobulin G2.";
RL Biochem. J. 121:217-225(1971).
RN [10]
RX DISULFIDE BONDS.
RX MEDLINE=69064124; PubMed=5782707;
RA Frangione B., Milstein C., Pink J.R.L.;
RT "Structural studies of immunoglobulin G.";
RL Nature 221:145-148(1969).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.ac.uk/announcements> or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; J00230; AAB59393.1; -.
DR PIR; A93906; G2HU.
DR HSSP; P01857; 1FC1.
DR Genew; HGNC:5526; IGHG2.
DR MIM; 147110; -.
DR GO; GO:0005624; C:membrane fraction; NAS.
DR GO; GO:0003823; F:antigen binding; TAS.
DR GO; GO:0006955; P:immune response; NAS.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003597; IG C1.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; Ig; 3.
DR SMART; SM00407; IgC1; 2.
DR PROSITE; PS00835; IG LIKE; 3.
DR PROSITE; PS00290; IG_MHC; 2.
KM Immunoglobulin domain; Immunoglobulin C region.
FT NON_TER 1
FT DOMAIN 1
FT 98
FT 99 110 CH1.
FT DOMAIN 111 219 HINGE.
FT 220 326 CH2.
FT DOMAIN 14 14 CH3.
FT DISULFID 27 83 INTERCHAIN (WITH A LIGHT CHAIN).
FT DISULFID 102 102 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 103 103 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 106 106 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 109 109 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 140 200
FT DISULFID 246 304
FT DISULFID 156 156
FT MOD_RES 326 326
FT VARIANT 60 60
FT CONFLICT 109 109
FT SEQ 326 AA; 35884 MW; 83108786878CF9C CRC64;
SQ
Query Match 32.7%; Score 1115.5; DB 1; Length 326;
Best Local Similarity 89.2%; Pred. No. 4,2e-64;
Matches 206; Conservative 8; Mismatches 4; Indels 13; Gaps 1;
OY 395 PCPAPBPKSCDKHTTCPELLGGPSVFLPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKF 454
DB 108 PCPAPP-----VAGPSVFLPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKF 154
OY 455 NWYVDGVVHNAKTKPREQVNSTFRVSVLTIVHQMVGKEYKCKVSNKGLPAPIEKT 514
DB 155 NWYVDGVVHNAKTKPREQVNSTFRVSVLTIVHQMVGKEYKCKVSNKGLPAPIEKT 214
OY 515 ISKAGQPREQVYTLTPPSDELTNQVSLTCLVKGYPSPDIADVEMSNQCPENNYKTPP 574
DB 215 ISKTKGPREQVYTLTPPSDELTNQVSLTCLVKGYPSPDIADVEMSNQCPENNYKTPP 274
OY 575 PVLDSDGSPFLYSKLTVDKSRMOQGNVPSGVMEALHNNHYTKSLSPG 625
DB 275 PVLDSDGSPFLYSKLTVDKSRMOQGNVPSGVMEALHNNHYTKSLSPG 325

```

RESULT 16
ID GC4_HUMAN STANDARD; PRT; 327 AA.
AC P01861;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig gamma-4 chain C region.
GN IGHG4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83157104; PubMed=6299662;
RA Ellison J.W., Buxbaum J.N., Hood L.E.;
RT "Nucleotide sequence of a human immunoglobulin C gamma 4 gene.";
RL DNA 1:11-18(1981).
RN [2]
RP SEQUENCE OF 1-30 AND 81-326.
RX MEDLINE=7020560; PubMed=4192699;
RA Pink J.R.L., Buttery S.H., de Vries G.M., Milstein C.;
RT "Human immunoglobulin subclasses. Partial amino acid sequence of the
RT constant region of a gamma 4 chain.";
RL Biochem. J. 117:33-47(1970).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: K01316; AAB59394.1; ALT_INT.
DR PIR: A90833; GAHU.
DR PDB: 1ADQ; 16-SEP-98.
DR Genew; HGNC:5528; IGHG4.
DR MIM; 147130; -.
DR GO: GO:0005624; C:membrane fraction; NAS.
DR GO: GO:0003823; F:antigen binding; TAS.
DR GO: GO:0006955; P:immune response; NAS.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003597; Ig_c1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig; 3.
DR SMART; SM00407; IgC1; 2.
DR PROSITE; PSS0835; IG_LIKE; 3.
DR PROSITE; PS00290; IG_MHC; 2.
KM Immunoglobulin domain; Immunoglobulin C region; 3D-structure.
FT NON_TER 1 1
FT DOMAIN 1 98 CH1.
FT DOMAIN 99 110 HINGE.
FT DOMAIN 111 220 CH2.
FT DOMAIN 221 327 CH3.
FT DISULFID 14 14 INTERCHAIN (WITH A LIGHT CHAIN).
FT DISULFID 27 83 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 106 106 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 109 109 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 141 201
FT DISULFID 247 305
SQ SEQUENCE 327 AA; 35940 MW; 3EDBD811EF208E7A CRC64;
Query Match 32.4%; Score 1105.5; DB 1; Length 327;
Best Local Similarity 89.6%; Pred. No. 1.8e-63;
Matches 207; Conservative 8; Mismatches 7; Indels 9; Gaps 2;
QY 395 PCPAPBPKSCDKHTCTCCELLGSPVFLPPKPKDTLMISRTPEVTCVVVDVSHEDPEVK 454
DB 105 PCP-----SC-----PAPEFLGSPVFLPPKPKDTLMISRTPEVTCVVVDVSHEDPEVK 155

```

```

QY 455 NMYVDGVEVHNAKTPREEQVNSTYRVVSVLTFLVTHQDMWINGKRYCKVSNKLPAPIEK 514
DB 156 NMYVDGVEVHNAKTPREEQVNSTYRVVSVLTFLVTHQDMWINGKRYCKVSNKLPAPIEK 215
QY 515 ISKAGQPREPQVYTLPPSRDELTKYNQVSLTCLVKGFPYSDIAVWESNGQPENNYKTP 574
DB 216 ISKAGQPREPQVYTLPPSQEEMTKNQVSLTCLVKGFPYSDIAVWESNGQPENNYKTP 275
QY 575 PVLDDSGFFLYSKLTVDKSRWQGNVPSCSYTMHMLNHNHYOKSLSLSPG 625
DB 276 PVLDDSGFFLYSKLTVDKSRWQGNVPSCSYTMHMLNHNHYOKSLSLSLG 326
RESULT 17
ID CD4_RAT STANDARD; PRT; 457 AA.
AC P05540;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
DE T4/Leu-3) (W3/25 antigen).
GN CD4.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=87175535; PubMed=3104900;
RA Clark S.J., Jefferies W.A., Barclay A.N., Gagnon J., Williams A.F.;
RT "Peptide and nucleotide sequences of rat CD4 (W3/25) antigen:
RT evidence for derivation from a structure with four
RT immunoglobulin-related domains.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:1649-1653(1987).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 210-393.
RX MEDLINE=93262437; PubMed=8493535;
RA Brady R.L., Dodson E.J., Dodson G.G., Lange G., Davis S.J.,
RA Williams A.F., Barclay A.N.;
RT "Crystal structure of domains 3 and 4 of rat CD4: relation to the
RT NH2-terminal domains.";
RL Science 260:979-983(1993).
CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
CC receptor interaction. May regulate T-cell activation.
CC -1- SUBUNIT: Associates with p56-lck (by similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M15768; AAA40901.1; -.
DR PIR: A27449; A27449.
DR PDB: 1CID; 15-JUL-93.
DR GlycoStatedB; P05540; -.
DR GO: GO:0042101; C:T-cell receptor complex; ISS.
DR GO: GO:0015026; F:coreceptor activity; ISS.
DR GO: GO:0042289; F:MHC class II protein binding; ISS.
DR GO: GO:0006955; P:immune response; ISS.
DR GO: GO:0045086; P:positive regulation of interleukin-2 biosyn. . .; ISS.
DR GO: GO:0030217; P:T-cell differentiation; ISS.
DR GO: GO:0045058; P:T-cell selection; ISS.
DR GO: GO:0007169; P:transmembrane receptor protein tyrosine kin. . .; ISS.
DR InterPro; IPR000973; CD4_TCSA.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003599; Ig.

```

DR Pfam; PF00047; Ig; 2.
 DR PRINTS; PR00692; CDATCANTIGEN.
 DR SMART; SMO0409; IG; 2.
 DR PROSITE; PS00835; IG-Like; 1.
 KM Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
 KM Immune response; Repeat; Signal; Lipoprotein; Palmitate; 3D-structure.
 FT SIGNAL 1 27
 FT CHAIN 28 457
 FT DOMAIN 28 394
 FT TRANSMEM 395 417
 FT DOMAIN 418 457
 FT DOMAIN 128 127
 FT DOMAIN 207 316
 FT DOMAIN 317 374
 FT CARBOHYD 186 186
 FT CARBOHYD 297 297
 FT CARBOHYD 392 392
 FT DISULFID 43 111
 FT DISULFID 158 187
 FT DISULFID 328 370
 FT LIPID 418 418
 FT LIPID 421 421
 FT STRAND 213 217
 FT TURN 218 219
 FT STRAND 222 225
 FT STRAND 232 243
 FT STRAND 252 258
 FT TURN 259 260
 FT STRAND 261 265
 FT STRAND 274 275
 FT STRAND 278 278
 FT TURN 280 280
 FT STRAND 282 285
 FT HELIX 290 292
 FT STRAND 294 301
 FT STRAND 306 319
 FT STRAND 325 331
 FT STRAND 338 344
 FT TURN 345 346
 FT STRAND 349 353
 FT STRAND 357 361
 FT STRAND 367 374
 FT TURN 375 376
 FT STRAND 377 385
 SQ SEQUENCE 457 AA; 51437 MW; 477BE157D30954C1 CRC64;
 Query Match 29.3%; Score 1000.5; DB 1; Length 457;
 Best Local Similarity 50.1%; Pred. No. 1.4e-56;
 Matches 213; Conservative 64; Mismatches 127; Indels 21; Gaps 6;
 QY 1 MNRGVPRHL--LVVLQALLPAATQGNKVVYLGKKGDVLTCTASOKKSIQFMKNSQ 58
 DB 1 MCRGSEFHLPLLLQLQSLVLTQGTIVLQKGGSAELPCSTSRSSAFMKSSDQ 60
 QY 59 IKIINGQSFLTKGSKLNDRAADSRSLMDQGNPFLIKNKIKIDSDPTICEVEDQKEV 118
 DB 61 KTIIGYKKKLITKSLLEYSRSPDSKKNMKGSPFLINIKLRMDSQTYVCELNKKEEV 120
 QY 119 QLVFGLTANSDFHLQGSITLTLES--PGSSPSVQCRSPRGKNIQGGKTLVSQLEIQ 177
 DB 121 ELWFRVYFNPCTRILQGSITLTLDNSNPKVSDPIECKHSSNIVKDSKAFSTHSLRIQ 180
 QY 178 DSGWTCTVLQNKQKVEPKIDI VVLAFOKASSIYKKEGEVFSFPLAFTVEKLTGSGE 237
 DB 181 DSGIMNCTVTLNOKKGSFDMKLSVIGFASTSITAVKSGEASERFPLNLGEEEL--QGE 238
 QY 238 LMQAERASSSKSWTTPDLKXKVEVVKVQDPKQMKKPLHLTLPLQALPQVAGSNTL 297
 DB 239 LRMAEKRPSSQSSTFSLKQKQVSVQKSTSNPKFQSLPTLQIPQVSLQFAGSNTL 298
 QY 298 TLAEAKTGKHOEVNLVVMRATOLQKN-LTCEVWGPRSPKMLSLKLENKAAVSKREK 356

DB 299 TLTLDD--RGILYORVNLVVMKVTPDSDNTLTCEVWGPTSPKMRLLIKQENQEARVSRQEK 356
 QY 357 PWTVLNPEAGWMOCLSDSOVLLESNIXILPTMTSTVPCAPRPKSCDHTHCELLGG 416
 DB 357 VIQVQAPAGVMOCLLSGEBVRKMSKIQVL-----SKGLNQTMPLAVLVGS 403
 QY 417 PSVPL 421
 DB 404 AFSL 408
 RESULT 18
 ID CD4 MOUSE STANDARD; PRT; 457 AA.
 AC P06332;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen T4/Leu-3) (T-cell differentiation antigen L3T4).
 GN CD4.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87018845; PubMed=3094146;
 RA Tourville B., Gorman S.D., Field E.H., Hunkapiller T., Parnes J.R.;
 RT "Isolation and sequence of L3T4 complementary DNA clones: expression in T cells and brain."
 RL Science 234:610-614(1986).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87115821; PubMed=3027575;
 RA Littman D.R., Gettner S.N.;
 RT "Unusual intron in the immunoglobulin domain of the newly isolated murine Cd4 (L3T4) gene."
 RL Nature 325:453-455(1987).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=88152875; PubMed=3326818;
 RA Parnes J.R., Hunkapiller T.;
 RT "L3T4 and the immunoglobulin gene superfamily: new relationships between the immune system and the nervous system."
 RL Immunol. Rev. 100:109-127(1987).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Brain;
 RX MEDLINE=88041159; PubMed=2823269;
 RA Gorman S.D., Tourville B., Parnes J.R.;
 RT "Structure of the mouse gene encoding CD4 and an unusual transcript in brain."
 RL Proc. Natl. Acad. Sci. U.S.A. 84:7644-7648(1987).
 RN [5]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98112780; PubMed=9445485;
 RA Aneari-Lari M.A., Oeltgen J.C., Schwartz S., Zhang Z., Muzny D.M., Lu J., Gortell J.H., Chinault A.C., Belmont J.W., Miller W., Gibbs R.A.;
 RT "Comparative sequence analysis of a gene-rich cluster at human chromosome 12p13 and its syntenic region in mouse chromosome 6."
 RL Genome Res. 8:29-40(1998).
 RN [6]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC STRAIN=C57BL/6J; TISSUE=Mammary gland;
 RX MEDLINE=22386257; PubMed=12477932;
 RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F., Datchenko L., Marsina K., Farmer A.A., Rubin G.M., Hong L.,

RA Strapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Ketterman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W.,
 RA Butcherfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences".
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [7]
 RN
 RP SEQUENCE OF 27-43.
 RX MEDLINE=6166694; PubMed=3082751;
 RA Classon B.J., Tsagaratos J., Kitzbaum L., Maddox J., McKay C.R.,
 RA Brandon M., McKenzie I.F.C., Walker I.D.,
 RT "The L314 antigen in mouse and the sheep equivalent are
 RT Immunoglobulin-like".
 RL Immunogenetics 23:129-132(1986).
 RN [8]
 RP DISULFIDE BONDS.
 RX MEDLINE=6233454; PubMed=3086886.
 RA Classon B.J., Tsagaratos J., McKenzie I.F.C., Walker I.D.,
 RT "Partial primary structure of the T4 antigens of mouse and sheep:
 RT assignment of intrachain disulfide bonds".
 RL Proc. Natl. Acad. Sci. U.S.A. 83:4499-4503(1986).
 CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
 CC receptor interaction. May regulate T-cell activation.
 CC -1- SUBUNIT: Associates with p56-lck (by similarity).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=P06332-1; Sequence=Displayed;
 CC Name=2; Synonyms=Brain-specific;
 CC IsoId=P06332-2; Sequence=VSP_002489;
 CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC
 DR EMBL; M36850; AAA39401.1; -;
 DR EMBL; M13816; AAA37267.1; -;
 DR EMBL; X04836; CAA28539.1; -;
 DR EMBL; M36851; AAA39402.1; -;
 DR EMBL; M17080; AAA37403.1; -;
 DR EMBL; M17078; AAA37403.1; JOINED.
 DR EMBL; M17079; AAA37403.1; JOINED.
 DR EMBL; AC002397; AAC36010.1; -;
 DR EMBL; BC039137; AAH39137.1; -;
 DR PIR; A02110; RWMST4.
 DR HSSP; P01730; 1MBR.
 DR MCD; MGI:88335; Cd4.
 DR GO; GO:0042101; C:T-cell receptor complex; ISS.
 DR GO; GO:0015026; F:coreceptor activity; ISS.
 DR GO; GO:0042289; F:MHC class II protein binding; ISS.
 DR GO; GO:0006955; P:immune response; ISS.
 DR GO; GO:0045086; P:positive regulation of interleukin-2 biosyn. . .; ISS.
 DR GO; GO:0030217; P:T-cell differentiation; ISS.
 DR GO; GO:0045058; P:T-cell selection; ISS.
 DR GO; GO:0007169; P:transmembrane receptor protein tyrosine kin. . .; ISS.
 DR InterPro: IPR000973; CD4 TCAG.
 DR InterPro: IPR007110; Ig-Like.

DR InterPro: IPR003596; Ig_V.
 DR Pfam: PR00047; Ig_2
 DR PRINTS; PR00062; CD4TCANTIGEN.
 DR SMART; SM00406; IGV; 1.
 DR PROSITE; PS00835; IG_LIKE; 1.
 KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
 KW Immune response; Repeat; Signal; Lipoprotein; Palmitate;
 KW Alternative splicing
 FT SIGNAL 1 26
 FT CHAIN 27 457
 FT DOMAIN 27 394
 FT TRANSMEM 395 417
 FT DOMAIN 418 457
 FT DOMAIN 27 128
 FT DOMAIN 129 207
 FT DOMAIN 208 317
 FT DOMAIN 318 374
 FT CARBOHYD 187 197
 FT CARBOHYD 298 298
 FT CARBOHYD 323 323
 FT CARBOHYD 392 392
 FT DISULFID 42 112
 FT DISULFID 159 188
 FT DISULFID 328 370
 FT LIPID 418 418
 FT LIPID 421 421
 FT VARSPLIC 1 240
 SQ SEQUENCE 457 AA; 51236 MW; 1BD4V527CB00D33 CRC64;
 Query Match 29.1%; Score 993.5; DB 1; Length 457;
 Best Local Similarity 50.9%; Pred. No. 3,9e-56;
 Matches 217; Conservative 66; Mismatches 120; Indels 23; Gaps 8;
 QY 1 MNRGVPFRH-LLVLTQALLPAPATGKNVLAQKGDVVELTCTASQKSIQPFMKNSNQI 59
 DB 1 MCRALSLRLRLLLLLQLSQLAVTQKTLVLKGEGBALPCSSQKKTITVFWKPSDOR 60
 QY 60 KILGNQG-SFLTKG--PSKLNDRADSRRLMNOGNFPLITIKULKIDSDSTYICEVDQGE 116
 DB 61 KILGQHGKGVLLIRGSSPSQF-DRFDSKKGAMKSGPPLINKLKMBDSQTYICELENRKE 119
 QY 117 EYOLVFGLTANSDFTLQSGSLTLTLFS--PGSSSVVOCRSRPNIGQKTLVSQGE 175
 DB 120 EYELMVFKXTFBPGSILQSGSLTLTLDSNKSVPNTLCKKIKKGVSGSKVLSNLR 179
 QY 176 LQDSGTWTCTVLQONKRVFKIDIVLAFQKASSIVYKKEGSEVERFSPLAFTVEKLTGS 235
 DB 180 VQDSDFWNTCTVTLQDKNMFMTLSVIGFQSTAITVYKSEGBSAEFSPLNFAEE--NGM 237
 QY 236 GELMWOAERASSSKSWITFDLKNKESYVKRVTDPKLQNGKTLPLHLTLPALPOYAGSG 295
 DB 238 GELMWAERKDSFPQWISPSINKKESYVQKSTKDLQKELTLPILTKIPVSYLQPSAGSG 297
 QY 296 NLTLEAKTGKLBHVLVWRATQLOKNLTCEVWGPTSPKLMJSLKLENKAKYSKKE 355
 DB 298 NLTLTLD--KGTLDHQLVNVKVAQALNLTLCVWGPSPKRLTLKQENGEARVSEGO 355
 QY 356 KPVYVNLNPEAGMOCCLSDSGVLLSESNIKVLPWTSTPVCAPAEKPSCKDTHTECLLG 415
 DB 356 KVVQVAVPETGLMOCCLSGDKVKNDSRIQVL-----SRGNQVTFLLACVIG 402
 QY 416 GPSVFL 421
 DB 403 GSGFPL 408
 RESULT 19
 GC3M MOUSE STANDARD; PRT; 398 AA.
 ID GC3M MOUSE
 AC P03987;
 DT 23-OCT-1986 (Rel. 02, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)


```

DR PDB; 1F11; 06-FEB-01.
DR PDB; 1F58; 29-DEC-99.
DR PDB; 1KCS; 24-JUL-02.
DR PDB; 1KCR; 11-MAY-02.
DR PDB; 25C9; 09-JUL-99.
DR MGD; MGI:96446; Igh-4.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003597; Ig_c1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig1; 2.
DR SMART; SM00407; IgC1; 2.
DR PROSITE; PS50835; IG_LIKE; 3.
DR PROSITE; PS00290; IG_MHC; 1.
KM Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
KW Alternative splicing; Transmembrane; 3D-structure.
FT NON_TER 1
FT DOMAIN 1 97 CH1.
FT DOMAIN 98 110 HINGE.
FT DOMAIN 111 217 CH2.
FT DOMAIN 218 324 CH3.
FT DISUFID 27 82
FT DISUFID 102 102 INTERCHAIN (WITH A LIGHT CHAIN).
FT DISUFID 104 104 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISUFID 107 107 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISUFID 109 109 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISUFID 138 198
FT CARBOHYD 174 174 N-LINKED (GLCNAC. . .).
FT DISUFID 244 302
FT TRANSMEM 340 357
FT DOMAIN 358 393 CYTOPLASMIC (POTENTIAL).
SQ SEQUENCE 393 AA; 43386 MW; 4CC88343B7A1CE27 CRC64;

Query Match 27.2%; Score 928.5; DB 1; Length 393;
Best Local Similarity 57.2%; Pred. No. 4,4e-52;
Matches 167; Conservative 51; Mismatches 47; Indels 27; Gaps 3;

QY 379 LLESNKIVLPFWSPVPCAP-EPKSGCDKH-----TCPELL 414
DB 56 VLDSDLTLSSSVTVSSPPSEVTTCNVAPASSTVKVKKIYPRDCCPCICTVPEV- 114
QY 415 GGSVFLFPKPKDMLISRTPEVTCVVVDVSHEDPEVKENMYVDGVEVNAKTPREDO 474
DB 115 --SSEFFPPKPDVLTITLPKVTCCVVVDISKDDPEVQSWVDVEVHTAQDQREED 172
QY 475 YNSTYRVSVLTVLHODMLNGKEYCKVSKKALPAPEKTIISAKGQPREPPVYTTPPSR 534
DB 173 FNSFFSVSELPIMHODMLNGKEYCKVNSAAPAPAEKTIISKTKGPKAPQVYITPPPK 232
QY 535 DELTKQVSLTCLVKGFPSPDIVEMESNGCPENNYKTPPVLDSSGPFYLSKLVNDS 594
DB 233 EQNAKDQVSLTCLMTTFDFPEDIVEMWOMQOPENNYKNTPIVNTNGSYFVYSKLVNDS 292
QY 595 RMOGVNFSVCSVNEALHNHYTKSLSPGLQDLETCAEADGELDGLWTT 646
DB 293 NMEAGNFTCSVLEHGLHNHTEKSLSHSPGLQDLETCAEADGELDGLWTT 344

RESULT 21
GC RABIT STANDARD; PRT; 323 AA.
AC POL870;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig gamma chain C region.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84030930; PubMed=6313520;
RA Bernstejn K.E., Alexander C.B., Mage R.G.;

```

```

RT "Nucleotide sequence of a rabbit IgG heavy chain from the recombinant
RT F-I haplotype.";
RN Immunogenetics 18:387-397(1983).
RL (2)
RP SEQUENCE OF 1-128.
RX MEDLINE=76135469; PubMed=1243651;
RA Pratt D.M., Mole L.E.;
RT "Sequence studies on the constant region of the Fd sections of rabbit
RT immunoglobulin G of different allotype.";
RN Biochem. J. 151:337-349(1975).
RL (3)
RP SEQUENCE OF 88-266 FROM N.A.
RX MEDLINE=83299917; PubMed=6193512;
RA Marcens C.L., Moore K.W., Steinmetz M., Hood L., Knight K.L.;
RT "Heavy chain genes of rabbit IgG: isolation of a cDNA encoding gamma
RT heavy chain and identification of two genomic C gamma genes.";
RN Proc. Natl. Acad. Sci. U.S.A. 79:6018-6022(1982).
RL (4)
RP SEQUENCE OF 132-161.
RX MEDLINE=70110015; PubMed=5461106;
RA Fruchter R.G., Jackson S.A., Mole L.E., Porter R.R.;
RT "Sequence studies of the Fd section of the heavy chain of rabbit
RT immunoglobulin G.";
RN Biochem. J. 116:249-259(1970).
RL (5)
RP SEQUENCE OF 129-131 AND 155-322.
RA Hill R.L., Lebovitz H.E., Fellows R.E. Jr., Delaney R.;
RT (in) Killander J. (eds.);
RL Gamma globulins, Nobel symp. 3, pp.109-127, Almqvist and Wiksell,
RL Stockholm (1967).
CC -1- MISCELLANEOUS; Ref.1 sequence has the D12 allelic marker,
CC 104-Thr, and Ref.5 the E14 marker. Ref.3 has the D11 and E15
CC markers and Ref.5 the E15 marker.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M16436; AAA31289.1; -
DR PIR; A91749; GHRB.
DR HSSP; P01857; 1FC1.
DR InterPro; IPR007110; Ig-1like.
DR InterPro; IPR003597; Ig_c1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig1; 2.
DR SMART; SM00407; IgC1; 2.
DR PROSITE; PS50835; IG_LIKE; 3.
DR PROSITE; PS00290; IG_MHC; 1.
KM Immunoglobulin domain; Immunoglobulin C region; Repeat.
FT NON_TER 1
FT DOMAIN 1 96
FT DOMAIN 114 213
FT DOMAIN 222 318
FT VARIANT 104 104 T->M (IN D11 MARKER).
FT VARIANT 185 185 T->A (IN E15 MARKER).
FT CONFLICT 48 48 N->E (IN REF. 2).
FT CONFLICT 71 71 V->VPV (IN REF. 2).
FT CONFLICT 144 144 Q->E (IN REF. 3 AND 4).
FT CONFLICT 173 173 N->D (IN REF. 5).
FT CONFLICT 187 187 Q->E (IN REF. 5).
FT CONFLICT 201 201 N->D (IN REF. 3 AND 5).
FT CONFLICT 218 218 Q->E (IN REF. 5).
FT CONFLICT 233 233 E->Q (IN REF. 5).
FT CONFLICT 246 246 N->D (IN REF. 5).
FT CONFLICT 256 256 E->G (IN REF. 5).
FT CONFLICT 260 260 N->D (IN REF. 5).
FT CONFLICT 266 266 N->D (IN REF. 5).
FT CONFLICT 280 280 Y->W (IN REF. 5).

```

FT CONFLICT 284 284 N -> S (IN REF. 5).
 SQ SEQUENCE 323 AA; 35404 MM; 69EBAA118D579A8B CRC64;
 Query Match 26.7%; Score 910; DB 1; Length 323;
 Best Local Similarity 65.7%; Pred. No. 5.1e-51;
 Matches 174; Conservative 31; Mismatches 44; Indels 16; Gaps 5;

QY 375 SGQVLEENIKYLPWSTPVP---PAPE-----PKSCDKTHTC--PELLGSPV 420
 DB 60 SGVLSLSVSVTS--SGPVTGCVNAHPATNTKDKTVAPTCSKP--TPEPELLGSPV 117
 QY 421 LPPPKKDTLMISRTPEVTCVVDVSHEDPEVKENMYDGVYVNAKTPREBYNSTYR 480
 DB 118 IFPPKPKDTLMISRTPEVTCVVDVSGDDPEVQFTWYINNBOVTRAPRLREQQFNSTIR 177
 QY 481 VVSUVTLVHOMLNKEKCKVSNKALPAPIEKTISKAKGQPREPQVYTLTPSRDELTKN 540
 DB 178 VVSTLPIRHQMLRKEKCKVHNKALPAPIEKTISKAKGQPLEKQVYTWGPPREELSSR 237
 QY 541 QVSLTCLVKGFPSPDIAYEWESNGQPENNYKTPPVLDSDGSFPLYSKLVTDKSRMOQN 600
 DB 238 VSLTCLMNGFPSPDISIVEMKNGKAEEDNYKTPPAVLDDSDSYFLYNKLSVPTSEMRGD 297
 QY 601 VFSCSVMEALHNHYTKSLSPG 625
 DB 298 VFTCSVMHEALHNHYTKSISRSPG 322

RESULT 22
 GCAM_MOUSE STANDARD; PRT; 399 AA.
 ID GCAM_MOUSE P01865;
 AC P01865;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Ig gamma-2A chain C region, membrane-bound form.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 NC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=82222190; PubMed=6283537;
 RA Yamawaki-Kataoka Y., Nakai S., Miyata T., Honjo T.;
 RT "Nucleotide sequences of gene segments encoding membrane domains of
 RT immunoglobulin gamma chains.";
 RL Proc. Natl. Acad. Sci. U.S.A. 79:2623-2627(1982).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=Membrane-bound;
 CC IsoId=P01865-1; Sequence=Displayed;
 CC Name=Secreted;
 CC IsoId=P01864-1; Sequence=External;
 CC Note=Probably the major isoform;
 CC -1- SIMILARITY: Contains 3 immunoglobulin-like domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isdb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; J00471; AB59661.1; ALT_INIT.
 CC PIR; A02154; G2MSAM.
 CC PDB; 1KB5; 06-APR-98.
 CC PDB; 1YEE; 15-OCT-97.
 CC MCD; M0196443; Igh-1.
 CC InterPro; IPR007110; Igh-1-like.
 CC InterPro; IPR003597; Igh_c1.
 CC InterPro; IPR003006; Igh_MHC.
 CC Pfam; PF00047; Igh; 2.

DR SMART; SMO0407; IGc1; 2.
 DR PROSITE; PS50835; IG_LIKE; 3.
 DR PROSITE; PS00290; IG_MHC; 1.
 KM Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
 KW Transmembrane; Alternative splicing; 3D-structure; Repeat.
 FT NON TRM 1
 FT DOMAIN 6 98
 FT DOMAIN 121 220
 FT DOMAIN 229 325
 FT DISULFID 15 15
 FT DISULFID 27 82
 FT DISULFID 107 107
 FT DISULFID 110 110
 FT DISULFID 112 112
 FT DISULFID 144 204
 FT DISULFID 250 308
 FT TRANSMEM 346 363
 FT DOMAIN 364 399
 FT CARBOHYD 180 180
 SQ SEQUENCE 399 AA; 44020 MM; 4C38138BPAED3FP0 CRC64;
 Query Match 26.4%; Score 902.5; DB 1; Length 399;
 Best Local Similarity 49.0%; Pred. No. 2e-50;
 Matches 180; Conservative 45; Mismatches 87; Indels 55; Gaps 5;

QY 285 PQLPQVAGSGNLTLEAKTGKHOEV-----NLVYMRATPOLQXNLTCEVWGPTSPKLM 339
 DB 34 PEPVTLTWNGSGLSGVTTPPAVLQSDLYTLSSVYTSSTPSQSLTCVNAHPAS---- 89
 QY 340 LSLTKENKAKVSKREKRPVWVLPNPEAGMOCLLSDSGQVLESNIKVLPTWSTPVP 399
 DB 90 -----STKVKD-----KIEPGRPTIKPCP-- 108
 QY 400 EPKSCDKTHTCPPELLGSPVFLPPPKKDTLMISRTPEVTCVVDVSHEDPEVKENMYD 459
 DB 109 -PCCK-----PAPNLGSPVFIPEPKIKDYLMSISPIYCVVDVSEDDPVDQISWVFN 163
 QY 460 GVEVHNKATKPREQVNSTRVVSVTLVTHODMNGEKVCKVSNKALPAPIEKTISKAK 519
 DB 164 NVEVHTAQOTQHRDYNSTLRVVSALPIQHDMWSSGEKCKVNNKOLPAPIEKTISKPK 223
 QY 520 GQPREPQVYTLTPSRDELTKNOVSLTCLVKGFPSPDIAYEWESNGQPENNYKTPPVLD 579
 DB 224 GSVAPRPQVYVLPPEEEMTKQVTLTCMVTDPMPEDIYVEMTNGKTELANKTEPVLDS 283
 QY 580 DGSFLLYSKLVTDKSRNQGNVSPGSVMHEALHNHYTKSLSPGLQDDETCAEADGE 639
 DB 284 DGSYFMYSKLRVEKKNVERNSSYSCVYHGLNHHHTTKSFSTRPGLDLDVCAEADGE 343
 QY 640 LDGLWTT 646
 DB 344 LDGLWTT 350

RESULT 23
 GC2_CAVPO STANDARD; PRT; 329 AA.
 ID GC2_CAVPO P01862;
 AC P01862;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Ig gamma-2 chain C region.
 OS Cavia porcellus (Guinea pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 NC Mammalia; Eutheria; Rodentia; Hystriocognathi; Cavidae; Cavia.
 CX NCBI_TaxId=10141;
 RN [1]
 RP SEQUENCE OF 1-3.
 RA Trischmann T.M.;
 RL Submitted (APR-1975) to the PIR data bank.
 RN [2]
 RP SEQUENCE OF 4-68.
 RX MEDLINE=71058471; PubMed=5538606;

RA Birnstein B.K., Huessein O.Z., Cebra J.J.;
 RT "Structure of heavy chain from strain 13 guinea pig
 immunoglobulin-G(2). 3. Amino acid sequence of the region around the
 half-cysteine joining heavy and light chains.";
 RL Biochemistry 10:18-25(1971).
 RN [3]
 RP SEQUENCE OF 69-133 AND 312-329.
 RX MEDLINE=71058466; PubMed=5538616;
 RA Turner K.J., Cebra J.J.;
 RT "Structure of heavy chain from strain 13 guinea pig
 immunoglobulin-G(2). II. Amino acid sequence of the carboxyl-terminal
 RT and hinge region cyanogen bromide fragments.";
 RL Biochemistry 10:9-17(1971).
 RN [4]
 RP SEQUENCE OF 134-226.
 RX MEDLINE=75036072; PubMed=4423665;
 RA Tracey D.E., Cebra J.J.;
 RT "Primary structure of the CH2 homology region from guinea pig IgG2
 RT antibodies.";
 RL Biochemistry 13:4796-4803(1974).
 RN [5]
 RP SEQUENCE OF 227-311.
 RX MEDLINE=75036073; PubMed=4609467;
 RA Trischmann T.M., Cebra J.J.;
 RT "Primary structure of the CH3 homology region from guinea pig IgG2
 RT antibodies.";
 RL Biochemistry 13:4804-4811(1974).
 RN [6]
 RP DISULFIDE BONDS.
 RX MEDLINE=71058474; PubMed=4922544;
 RA Oliveira B., Lamm M.E.;
 RT "Interchain disulfide bridges of guinea pig gamma-2-immunoglobulin.";
 RL Biochemistry 10:26-31(1971).
 CC -1- MISCELLANEOUS: This chain was isolated from pooled serum of strain
 CC 13 inbred guinea pigs.
 CC PIR: A94553; G2GP.
 DR HSSP: P01842; 7PAB.
 DR InterPro: IPR007110; Ig-like.
 DR InterPro: IPR003597; Ig_G1.
 DR InterPro: IPR003006; Ig_MHC.
 DR Pfam: PF00047; Ig_2.
 DR SMART: SM00407; IgG1; 2.
 DR PROSITE: PS50835; IG_LIKE; 3.
 DR PROSITE: PS00290; IG_MHC; 1.
 KM Immunoglobulin domain; Immunoglobulin C region; Glycoprotein.
 FT NON_TER 1 1
 FT DISULFID 16 16 INTERCHAIN (WITH A LIGHT CHAIN).
 FT DISULFID 28 79
 FT DISULFID 105 105 INTERCHAIN (WITH A HEAVY CHAIN).
 FT DISULFID 107 107 INTERCHAIN (WITH A HEAVY CHAIN).
 FT DISULFID 110 110 INTERCHAIN (WITH A HEAVY CHAIN).
 FT DISULFID 142 202
 FT CARBOHYD 178 178 N-LINKED (GLCNAC. . .).
 FT DISULFID 248 308
 SQ SEQUENCE 329 AA; 36074 MM; SD231B7164D1FBA9 CRC64;
 Query Match 25.8%; Score 879.5; DB 1; Length 329;
 Best Local Similarity 69.2%; Pred. No. 4.6e-49;
 Matches 164; Conservative 27; Mismatch 39; Indels 7; Gaps 3;
 QY 393 PVPCPAPKPSKCDTHTC--PELLGGSVFLPPKPKDPTLMISTPPTCVVDVSHEDP 450
 DB 96 PIRIPZBPTCTCFK---CPPPENIGGSVFIFFPKPKDPTLMISTPPTCVVDVSHEDP 152
 QY 451 EVKENMYVDGVEVNAKTKPREQYNSTYRVSVLTVLHODMLNGEKYCKVSNKALPAP 510
 DB 153 EVQFTWVDVNDKPVNAETKPRVEQYNTTFRVESLPIQHODMLNGKGFCKVYNNKALPAP 212
 QY 511 IEKTSIRAKQPREPQYITLPPSRDELTKNOVSLTLCVKGFPESDIAVENESGQP--EN 568
 DB 213 IEKTSIRKGAIRMPDYVTLPPSRDELTSKSVSYTCLINFFPADIHVEWASNVVSEK 272
 QY 569 NYKTPPVLDSDGSFLYSKLTVDKSRNQGNVSCSVMEHALNHTQKSLSLSPG 625

DB 273 EKXNPPIEDADGSLFLYSKLTVDKSRNQGNVSCSVMEHALNHTQKSLSLSPG 329
 RESULT 24
 GCNM MOUSE
 ID GCNM_MOUSE STANDARD; PRT; 405 AA.
 AC P01867;
 DT 21-UTR-1986 (Rel. 01, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Ig gamma-2B chain C region, membrane-bound form.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE OF 335-405 FROM N.A.
 RX MEDLINE=82222190; PubMed=6283537;
 RA Yamawaki-Kataoka Y., Nakai S., Miyata T., Honjo T.;
 RT "Nucleotide sequences of gene segments encoding membrane domains of
 RT immunoglobulin gamma chains.";
 RL Proc. Natl. Acad. Sci. U.S.A. 79:2623-2627(1982).
 RN [2]
 RP SEQUENCE OF 335-378 FROM N.A.
 RX MEDLINE=82115285; PubMed=6799207;
 RA Rogers J., Choi E., Souza L., Carter C., Word C.J., Kuehl M.,
 RA Eisenberg D., Wall R.;
 RT "Gene segments encoding transmembrane carboxyl terminal of
 RT immunoglobulin gamma chains.";
 RL Cell 26:19-27(1981).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=Membrane-bound;
 CC IsoId=P01867-1; Sequence=Displayed;
 CC Name=Secreted;
 CC IsoId=P01866-1; Sequence=External;
 CC Note=May be the major isoform;
 CC -1- PTM: DISULFIDE BONDS BY HOMOLOG WITH OTHER IG GAMMA CHAINS.
 CC -1- MISCELLANEOUS: The sequence of residues 1-335 is assumed to be
 CC identical with the corresponding region of the secreted form.
 CC -1- MISCELLANEOUS: The a allele sequence is shown.
 CC -1- SIMILARITY: Contains 3 immunoglobulin-like domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: J00462; AAB59659.1; ALT_INIT.
 DR PIR: C02154; G2MSBM.
 DR PDB: 1C1C; 11-MAR-03.
 DR MGI: 96445; IgH-3.
 DR InterPro: IPR007110; Ig-like.
 DR InterPro: IPR003597; Ig_G1.
 DR InterPro: IPR003006; Ig_MHC.
 DR Pfam: PF00047; Ig_3.
 DR SMART: SM00407; IgG1; 2.
 DR PROSITE: PS50835; IG_LIKE; 3.
 DR PROSITE: PS00290; IG_MHC; 1.
 KM Immunoglobulin domain; Immunoglobulin C region; Transmembrane;
 KM Alternative splicing; 3D-structure; Repeat.
 FT NON_TER 1 1
 FT DOMAIN 6 98 IG-LIKE 1.
 FT DOMAIN 127 226 IG-LIKE 2.
 FT DOMAIN 235 331 IG-LIKE 3.
 FT DISULFID 15 15 INTERCHAIN (WITH A LIGHT CHAIN).
 FT DISULFID 27 82 INTERCHAIN (WITH A LIGHT CHAIN).
 FT DISULFID 109 109 INTERCHAIN (WITH A HEAVY CHAIN).
 FT DISULFID 112 112 INTERCHAIN (WITH A HEAVY CHAIN).


```

FT DISULFID 111 111 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 140 200
FT DISULFID 246 304
FT CARBOHYD 176 176 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 326 AA; 35946 MW; 013BAB45EF49B9DA CRC64;

Query Match 24.3%; Score 829; DB 1; Length 326;
Best Local Similarity 57.3%; Pred. No. 7.4e-46;
Matches 153; Conservative 42; Mismatches 56; Indels 16; Gaps 4;

QY 375 SCQVLLSNIKIV-LPTW-STVPDPCAPPEPKSCDKT--HTCPPELLGG-----PS 418
DB 59 SGLYLTLSSTVTSVPSQTVTCNVANHPASSSTKVKKIVPRNCGGDCRKCICITGSEVSS 118
QY 419 VFLEPPPKQDTLMISRTPEVTCVVVDVSHEDPEVKFNWYDGVENNAKTKPEEQYNSI 478
DB 119 VFLEPPPKQDTLTITLTPKTCVAVVDISODDPEVHFSWFDVDEVTAAQTRPEEQFNST 178
QY 479 YRVVSVLTVTHQOMLNQKEVKCKVSKNKAAPLPIEKITSKAKGQPREPOVYTLPPSRDEL 538
DB 179 FRVSVSELPILHQMILNKRTRCKVTSAAFPSPLEKITSKEPKRTQVPHVYTNMSPTEEMT 238
QY 539 KNOVSLTCLVKGFPYPSDIAVEMESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRMQQ 598
DB 239 QNEVSTICWVKGFPYPPDIYVEMQNGQPQENYKNTTPMTDSDSYFLYSKLVNKKKEMQQ 298
QY 599 GNVFSCGVMEALHNHYTQKSLSLSPG 625
DB 299 GNVFTCSVLHLEGLHNHTERSLSHSPG 325

RESULT 27
GCB_RAT ID GCB_RAT STANDARD; PRT; 333 AA.
AC P20761;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig gamma-2B chain C region.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89232738; PubMed=3149946;
RA Brueggemann M.;
RT "Evolution of the rat immunoglobulin gamma heavy-chain gene family.";
RL Gene 74:473-482(1988).
CC -1- SIMILARITY: Contains 3 immunoglobulin-like domains.
DB PIR; PS0018; PS0018.
DB HSSP; P01842; 7FAB.
DR InterPro; IPR007110; Ig-1like.
DR InterPro; IPR003597; Ig_c1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig_3.
DR SMART; SM00407; IgC1; 2.
DR PROSITE; PS50835; Ig_LIKE; 3.
DR PROSITE; PS00290; Ig_MHC; 1.
KW Immunoglobulin domain; Immunoglobulin C region; Repeat.
FT NON TER 1 1
FT DOMAIN 1 96 IG-LIKE 1.
FT DOMAIN 124 223 IG-LIKE 2.
FT DOMAIN 232 328 IG-LIKE 3.
FT DISULFID 15 15 INTERCHAIN (WITH A LIGHT CHAIN).
FT DISULFID 27 80 INTERCHAIN (WITH A LIGHT CHAIN).
FT DISULFID 106 106 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 109 109 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 112 112 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 115 115 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 147 207 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 253 311
SQ SEQUENCE 333 AA; 36497 MW; 55F8B64D48D460A6 CRC64;

```

```

Query Match 24.2%; Score 826.5; DB 1; Length 333;
Best Local Similarity 57.1%; Pred. No. 1.1e-45;
Matches 157; Conservative 35; Mismatches 58; Indels 25; Gaps 4;

QY 375 SCQVLLSNIKIV-LPTW-STVPDPCAPPEPKS-----CDKHTTC-- 410
DB 59 SGLYLTLSSTV-TSSVPSQTVTCNVANHPASSSTKVKKIVPRNCGGDCRKCICITGSEVSS 117
QY 411 PELGGSVYFLPPPKQDTLMISRTPEVTCVVVDVSHEDPEVKFNWYDGVENNAKTKP 470
DB 118 PELGGSVYFLPPPKQDTLTITLTPKTCVAVVDISODDPEVHFSWFDVDEVTAAQTRPEEQ 177
QY 471 REEOVNSTYRVVSVLTVTHQOMLNQKEVKCKVSKNKAAPLPIEKITSKAKGQPREPOVYTL 530
DB 178 REEOVNSTYRVVSVLPIDQDMNSGKEFKCKVNNKALPSPIEKITSKPKGLVKKPQVYVM 237
QY 531 PPSRDELTKNQVSLTCLVKGFPYPSDIAVEMESNGQPENNYKTTPPVLDSDGSFFLYSKLT 590
DB 238 GPPTGLTEQTVSLTCLTSGFLPNDIGVEMTNSGHIKKYKNTPEVMDSDGSFFMYSKLN 297
QY 591 YDKSRMQQGNVSCVMEALHNHYTQKSLSLSPG 625
DB 298 VERSRWDSRAPVCSVMEGLHNHVEKSIKSRPG 332

RESULT 28
GCC_RAT ID GCC_RAT STANDARD; PRT; 329 AA.
AC P20762;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig gamma-2C chain C region.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88166903; PubMed=3127222;
RA Brueggemann M.; Delmasstro-Galife P.; Waldmann H.; Calabi F.;
RT "Sequence of a rat immunoglobulin gamma 2c heavy chain constant
region cDNA: extensive homology to mouse gamma 3."
RL Eur. J. Immunol. 18:317-319(1988).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1sb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DB EMBL; X07189; CAJ0169.1; -.
DB PIR; S00847; S00847.
DR HSSP; P01842; 7FAB.
DR InterPro; IPR007110; Ig-1like.
DR InterPro; IPR003597; Ig_c1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig_2.
DR SMART; SM00407; IgC1; 2.
DR PROSITE; PS50835; Ig_LIKE; 3.
DR PROSITE; PS00290; Ig_MHC; 1.
KW Immunoglobulin domain; Immunoglobulin C region.
FT NON TER 1 1
FT DOMAIN 1 97 CH1.
FT DOMAIN 98 113 HINGE.
FT DOMAIN 114 222 CH2.
FT DOMAIN 223 329 CH3.
FT DISULFID 15 15 INTERCHAIN (WITH A LIGHT CHAIN).
FT DISULFID 27 82 INTERCHAIN (WITH A LIGHT CHAIN).
FT DISULFID 111 111 INTERCHAIN (WITH A HEAVY CHAIN).

```

FT	DISULFID	113	113	INTERCHAIN (WITH A HEAVY CHAIN).
FT	DISULFID	143	203	
FT	DISULFID	249	307	
SO	SEQUENCE	329 AA,	36571 MW,	5 FCD7B7933850773 CRC64;
Query Match		24.1%;	Score 822.5;	DB 1; Length 329;
Best Local Similarity		57.0%;	Pred. No. 2e-45;	
Matches	154;	Conservative	47;	Mismatches 50; Indels 19; Gaps 4
Oy	375	SGQVLLSNIKV-LPTWST-PVPC----	DA-----	PEPKSCDKHTTCELG 415
Db	59	SGLYTLSSSVTPSSSTSSQVATCSVAHDAKTSNLIKRIEPRRKRPRPTDICSODNIG	118	
Oy	416	GPSVLLFPKKPDITLMISTPEVTCVVDVSHDEPVKRWYVDGVEVNAKTRPEEQY	475	
Db	119	RPSVEIFPEKPDIMITLTLPKTCVVDVSEBEPVQSPFADNVRVFPATQTPHEEQ	178	
Oy	476	NSTYVAVSLTVLHDDWMLNGKEKVCNNKALPAPIEKTISKAGQPREPOVYTLPPSRD	535	
Db	179	NGTFVAVSLTHQHODMWSGKEFKCKVNNKDLPSLEKTSIKRGAKRTPOVYTLPPRE	238	
Oy	536	ELTKQVSLTCLVKGFPSPDIAVEMESQCPENNYKTTTPVLDSDGSFFLYSLKLYDKSR	595	
Db	229	QMSKKVSLTCLVNTSPSPASISVEWERNELDEDYKNTLPVLDSDGSFFLYSLKLYDTS	298	
Oy	596	WQGVNFGSCVNHGALHNHYTKSLSLSPG	625	
Db	299	WMRGDIYTCSSVNHGALHNHHTQKNLSRSPG	328	
RESULT 29				
GCI_MOUSE				
ID_GCI_MOUSE	STANDARD;	PRT;	324 AA.	
AC	P01868;			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	21-JUL-1986 (Rel. 01, Last sequence update)			
DT	15-MAR-2004 (Rel. 43, Last annotation update)			
DE	19 gamma-1 chain C region secreted form.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxId=10090;			
RN	[1]			
RN	SEQUENCE FROM N.A.			
RA	MEDLINE=80045036; PubMed=115593;			
RA	Honjo T., Obata M., Yamawaki-Kataoka Y., Kataoka T., Kawakami T.,			
RA	Takahashi N., Mano Y.,			
RT	"Cloning and complete nucleotide sequence of mouse immunoglobulin			
RT	gamma 1 chain gene.";			
RL	Cell 18:559-568(1979).			
RN	[2]			
RN	SEQUENCE OF 76-324 FROM N.A. (MYELOMA PROTEIN MOPC 31C).			
RX	MEDLINE=80202559; PubMed=6769752;			
RA	Obata M., Yamawaki-Kataoka Y., Takahashi N., Kataoka T., Shimizu A.,			
RA	Mano Y., Seidman J.G., Peterlin B.M., Leder P., Honjo T.,			
RT	"Immunoglobulin gamma 1 heavy chain gene: structural gene sequences			
RT	cloned in a bacterial plasmid.";			
RL	Gene 9:87-97(1980).			
RN	[3]			
RN	SEQUENCE OF 70-322 FROM N.A. (MYELOMA PROTEIN MOPC 21).			
RX	MEDLINE=80012837; PubMed=113776;			
RA	Rogers J., Clarke P., Salser W.,			
RT	"Sequence analysis of cloned cDNA encoding part of an immunoglobulin			
RT	heavy chain.";			
RL	Nucleic Acids Res. 6:3305-3321(1979).			
RN	[4]			
RN	SEQUENCE (MYELOMA PROTEIN MOPC 21).			
RX	MEDLINE=78242288; PubMed=98524;			
RA	Adeuqbo K.,			
RT	"Evolution of immunoglobulin subclasses. Primary structure of a			
RT	murine myeloma gamma chain.";			
RL	J. Biol. Chem. 253:6068-6075(1978).			
RN	[5]			

RP	DISULFIDE BONDS (MOPC 21).
RX	MEDLINE=73008889; PubMed=5073237;
RA	Svasti J., Milstein C.;
RT	"The disulfide bridges of a mouse immunoglobulin G1 protein.";
RL	Biochem. J. 126:837-850(1972).
CC	-I- SUBCELLULAR LOCATION: Secreted.
CC	-I- ALTERNATIVE PRODUCTS:
CC	Event=Alternative splicing; Named isoforms=2;
CC	Name=Secreted;
CC	IsoId=P01868-1; Sequence=Displayed;
CC	Note=May be the major isoform;
CC	Name=Membrane-bound;
CC	IsoId=P01869-1; Sequence=External;
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; V00793; CAA24172.1; -
DR	EMBL; V00793; CAA24173.1; -
DR	EMBL; V00793; CAA24174.1; -
DR	EMBL; V00793; CAA24175.1; -
DR	EMBL; V00795; CAA24176.1; -
DR	PIR; A02159; GIMS.
DR	GlycoSuiteDB; P01868; -
DR	MGI; MGI:96446; Igh-4.
DR	InterPro; IPRO07110; Ig-like.
DR	InterPro; IPRO03597; Ig_c1.
DR	InterPro; IPRO03006; Ig_mHC.
DR	Pfam; PF00047; Ig; 2.
DR	SMART; SMO0407; IGc1; 2.
DR	PROSITE; PS50835; IG_LIKE; 3.
DR	PROSITE; PS00290; IG_MHC; 1.
KW	Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
KV	Immunoglobulin chain; Alternative splicing.
KW	Non TER
FT	DOMAIN 1 97 CH1.
FT	DOMAIN 98 110 HINGE.
FT	DOMAIN 111 217 CH2.
FT	DOMAIN 218 324 CH3.
FT	DISULFID 27 82 INTERCHAIN (WITH A LIGHT CHAIN).
FT	DISULFID 102 102 INTERCHAIN (WITH A HEAVY CHAIN).
FT	DISULFID 104 104 INTERCHAIN (WITH A HEAVY CHAIN).
FT	DISULFID 107 107 INTERCHAIN (WITH A HEAVY CHAIN).
FT	DISULFID 109 109 INTERCHAIN (WITH A HEAVY CHAIN).
FT	DISULFID 138 198 N-LINKED (GLCNAC. .).
FT	CARBOHYD 174 174 /FTid-CAR_000055.
FT	DISULFID 244 302 REMOVED POST-TRANSLATIONALLY.
FT	MOD_RSS 324 324 N -> D (IN REF. 3).
FT	CONFLICT 276 276 N -> D (IN REF. 3).
FT	CONFLICT 278 278 N -> D (IN REF. 3).
SO	SEQUENCE 324 AA; 35704 MW; A338812P3DI2C93 CRC64;
Qy	Query Match 23.9%; Score 814.5; DB 1; Length 324;
Db	Best Local Similarity 53.9%; Pred. No. 6,2e-45;
Db	Matches 146; Conservative 51; Mismatches 47; Indels 27; Gaps 3;
Oy	379 LLESNIKVLPTWSIPVPCAP-EPKSGDKH-----TCPELL 414
Db	56 VLQSDLTLLSSSVTPSPPSSEIVCTCVNHPASTKVDDKKIVPRDCGCKPCICTYREV- 114
Oy	415 GGPSEVFPEPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNMYVDGVENVNAATKPREEQ 474
Db	115 --SSVFIFPPKPKDVLTITLPKYTCVVYDISKDDEPVQSWVFDVETATQTQPREBQ 172
Oy	475 YNSTYRVASVLTVLHODMLNGEKYCKVSKNALPAPIEKTISKAKGPREEPYVTLPSPR 534
Db	173 FNSGFRSVESELPFHODMLNGEKFKCVNSAAPAPAEKTIISTKGKPKAPOVYVTLPPK 232

```

OY 535 DELTKNOVSLTCLVKGFPSDIAVEMESNGCPENNYKTPPYLDSGSEFLYSKLTJYDKS 594
DB 233 EQAKDKVSLTCLITTFDPEDITYEMQNGQPAENYKNTOPIMNTGNSYFVYSKLVNOKS 292
OY 595 RMOQGVNFGSCSVMEALHNHYTKSLSLSPG 625
DB 293 NMEAGNFTCSVLHNEGJHNHTKSLSHSPG 323

RESULT 30
GCA_MOUSE STANDARD; PRT; 330 AA.
ID GCA_MOUSE
AC P01863;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig gamma-2A chain C region, A allele.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=81076554; PubMed=6777755;
RA Sikorav J.-L., Auffray C., Rougeon F.;
RT "Structure of the constant and 3' untranslated regions of the murine
RT Balb/c gamma 2a heavy chain messenger RNA.";
RL Nucleic Acids Res. 8:3143-3155 (1980).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=81198976; PubMed=6262729;
RA Yamawaki-Kataoka Y., Miyata T., Honjo T.;
RT "The complete nucleotide sequence of mouse immunoglobulin gamma 2a gene
RT and evolution of heavy chain genes: further evidence for intervening
RT sequence-mediated domain transfer.";
RL Nucleic Acids Res. 9:1365-1381 (1981).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=81223894; PubMed=6787604;
RA Olo R., Auffray C., Morchamps C., Rougeon F.;
RT "Comparison of mouse immunoglobulin gamma 2a and gamma 2b chain genes
RT suggests that exons can be exchanged between genes in a multigenic
RT family.";
RL Proc. Natl. Acad. Sci. U.S.A. 78:2442-2446 (1981).
RN [4]
RP MYELOMA PROTEIN MOPC 173.
RX MEDLINE=74175517; PubMed=4831970;
RA Bourgois A., Fougereau M., Rocca-Serra J.;
RT "Determination of the primary structure of a mouse IgG2a
RT immunoglobulin amino-acid sequence of the Fc fragment. Implications
RT for the evolution of immunoglobulin structure and function.";
RL Eur. J. Biochem. 43:423-435 (1974).
RN [5]
RP DISULFIDE BONDS.
RX MEDLINE=73056887; PubMed=4565406;
RA de Preval C., Fougereau M.;
RT "Determination of the primary structure of a mouse gamma G2a
RT immunoglobulin. Identification of the disulfide bridges.";
RL Eur. J. Biochem. 30:452-462 (1972).
CC -1- SIMILARITY: Contains 3 immunoglobulin-like domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC CC
CC EMBL; V00798; CAA24178.1; -.
CC PIR; A02152; GZMSA.
CC PDB; 1E4W; 12-JUL-01.

```

```

DR PDB; 1E4X; 12-JUL-01.
DR PDB; 1MNJ; 06-MAY-99.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003557; Ig_C1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig_2.
DR SMART; SM00407; IgC1; 2.
DR PROSITE; PS50835; IG_LIKE; 3.
DR PROSITE; PS00290; IG_MHC; 1.
KM Immunoglobulin domain; Immunoglobulin C region; 3D-structure; Repeat.
FT NON_TER 1 1
FT DOMAIN 6 98 IG-LIKE 1.
FT DOMAIN 121 220 IG-LIKE 2.
FT DOMAIN 229 325 IG-LIKE 3.
FT DISULFID 15 15 INTERCHAIN (WITH A LIGHT CHAIN).
FT DISULFID 27 82 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 107 107 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 110 110 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 112 112 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 144 204
FT DISULFID 250 308
FT MOD_RES 330 330 REMOVED POST-TRANSLATIONALLY.
SQ SEQUENCE 330 AA; 36389 MW; B84361C54A5A6864 CRC64;

Query Match 23.5%; Score 801.5; DB 1; Length 330;
Best Local Similarity 46.8%; Pred. 4.3e-44;
Matches 162; Conservative 44; Mismatches 85; Indels 55; Gaps 5;

OY 285 PQALPOYAGSGNLTALAEKTKLHQEV-----NLVVMRATOLQKNLTCEWGPSPKLM 339
DB 34 PEPVTLTNMNGSLSSGVHTFPVAVLQSDLYTLSSVTSVTSTWPSQGITCVAHAPAS----- 89
OY 340 LSLKLENKAKYKSKRKPWVNLNPEAGMQLSDSGQVLBSNITVLPWSPVPCPAP 339
DB 90 -----STKXDK-----XIEPGPIKPCP-- 108
OY 400 EPKSCDKHTCPPELGGPSVFLPPPKKDTLMISRTPEYTCVVDVSHDEPKFMVYD 459
DB 109 -PKC-----PAPNLGGPSVFLPPPKIKDVLMSIPYTCVVDVSEDDPDVQISFVN 163
OY 460 GVEVHNAKTRPREOYNSTYRVSVLTVLHODMLNKEKXKYNKALPAPIKTKISKAK 519
DB 164 NVEVHTAQIQTHREDYNSTLRVVSALPIQHDMWSKKEKCKVNNKDLPAPIRTISKPK 223
OY 520 GQPREQVYTLPPSRDELTKNQVSLTCLVKGFPSDIAVEMESNGCPENNYKTPPYLDS 579
DB 224 GSVRAQVYVLPPEPEEMTKQVTLTCVTDPEPDIYVEWTNGKTELNKXTEPELDS 283
OY 580 DSGSFYSKLTVDKSRMOQGVNFGSCSVMEALHNHYTKSLSLSPG 625
DB 284 DGSYFVYSKLTVRKKNWVERNSYSGSVHNEGJHNHTTTSFKTPG 329

RESULT 31
GCA_RAT STANDARD; PRT; 322 AA.
ID GCA_RAT
AC P20760;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig gamma-2A chain C region.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=892232738; PubMed=3149946;
RA Bruggemann M.;
RT "Evolution of the rat immunoglobulin gamma heavy-chain gene family.";
RL Gene 74:473-482 (1988).
CC -1- SIMILARITY: Contains 3 immunoglobulin-like domains.
CC CC

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: M13804; AAA4376.1; ALT_INTL.
 DR PIR: PS0019; PS0019.
 DR HSSP: P01842; 7FAB.
 DR InterPro: IPR007110; Ig-like.
 DR InterPro: IPR003597; Ig cl.
 DR InterPro: IPR003006; Ig_MHC.
 DR Pfam: PF00047; Ig; 2.
 DR SMART: SM00407; Igcl; 2.
 DR PROSITE: PS50835; IG-LIKE; 3.
 DR PROSITE: PS00290; IG_MHC; 1.
 KM Immunoglobulin domain; Immunoglobulin C region; Repeat.
 FT NON TER 1 1
 FT DOMAIN 6 98 IG-LIKE 1.
 FT DOMAIN 115 212 IG-LIKE 2.
 FT DOMAIN 221 317 IG-LIKE 3.
 FT DISULFID 27 82
 FT DISULFID 102 102 INTERCHAIN (WITH A HEAVY CHAIN).
 FT DISULFID 105 105 INTERCHAIN (WITH A HEAVY CHAIN).
 FT DISULFID 107 107 INTERCHAIN (WITH A HEAVY CHAIN).
 FT DISULFID 136 196
 FT DISULFID 242 300
 FT CARBOHYD 172 172
 SQ SEQUENCE 322 AA; 35186 MW; E8EA136A9DE01EDB CRC64; N-LINKED (GLCNAC. . .) (POTENTIAL).

Query Match 23.2%; Score 792; DB 1; Length 322;
 Best Local Similarity 56.4%; Pred. No. 1.7e-43;
 Matches 150; Conservative 39; Mismatches 59; Indels 18; Gaps 4;

QY 375 SGQVLLSNIKY-LPTMST-VVPCPAPKSCDKTH-----TCPELLGGPSV 419
 DB 59 SGLTTLTSSVTVPSSTWSSQAVTCNVAFPASTKVDKIVRECNPCGCTGSEV---SSV 115
 QY 420 FLPPPKPDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVNAKTKRREQDYNSTY 479
 DB 116 FIFPKTKDVLTLTLTPKVCVVVDISGNDEVRFSMWIDVEVHTAQTAPAEKQSNSTL 175
 QY 480 RVSVVLTVLHODMNGEKYCKVKSNKALPAPIEKTISAKQCPPEPVYTLPPSRDELTK 539
 DB 176 RSVSELPVHWDMLNGKTKFKCKVNSGAFPAPIEKSISKEGTPRGQVYTWAPPKIEWTQ 235
 QY 540 NOVSLTCLVNGFYPSDIAVEMESNGOPENNYKTPPVLDSCGFIVSKLTVDSRQOG 599
 DB 236 SQVSIITCVKGFPPDITTEMKMGQPOENYKNTPTMTDTSYFLSKLVNKKETWQOG 295
 QY 600 NVEFSCVWHEALAHNYTQKSLSLSPG 625
 DB 296 NTFCTSVLHBEGLHHNHTKSLSHSPG 321

RESULT 32
 GCAB_MOUSE STANDARD; PRT; 335 AA.
 AC P01864;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Ig gamma-2A chain C region secreted form (B allele).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6;
 RX MEDLINE=82037861; PubMed=6170065;

RA Schreier P.H., Bothwell A.L.M., Mueller-Hill B., Baltimore D.;
 RT "Multiple differences between the nucleic acid sequences of the
 RT Igg2a and Igg2b alleles of the mouse."
 RL Proc. Natl. Acad. Sci. U.S.A. 78:4495-4499(1981).
 RN [2]
 RN SEQUENCE.
 RX MEDLINE=82037777; PubMed=6794027;
 RA Dognin M.J., Lauwereys M., Stroberg A.D.;
 RT "Multiple amino acid substitutions between murine gamma 2a heavy
 RT chain Fc regions of Ig1a and Ig1b allotypic forms."
 RL Proc. Natl. Acad. Sci. U.S.A. 78:4031-4035(1981).
 CC -1- SUBCELLULAR LOCATION: Secreted (Potential).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=Secreted;
 CC IsoId=P01864-1; Sequence=displayed;
 CC Note=Probably the major isoform;
 CC Name=Membrane-bound;
 CC IsoId=P01865-1; Sequence=External;
 CC -1- MISCELLANEOUS: The sequence differs from that of the a allele,
 CC from BALB/c mice, at 15% of the positions.
 CC -1- SIMILARITY: Contains 3 immunoglobulin-like domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: J00479; -; NOT_ANNOTATED_CDS.
 DR PIR: A02153; GZMSAB.
 DR PDB: 1BOG; 23-MAR-99.
 DR PDB: 1HH6; 26-JAN-01.
 DR PDB: 1HH9; 24-JUL-03.
 DR PDB: 1HI6; 08-FEB-01.
 DR InterPro: IPR007110; Ig-like.
 DR InterPro: IPR003597; Ig cl.
 DR InterPro: IPR003006; Ig_MHC.
 DR Pfam: PF00047; Ig; 2.
 DR SMART: SM00407; Igcl; 2.
 DR PROSITE: PS50835; IG-LIKE; 3.
 DR PROSITE: PS00290; IG_MHC; 1.
 KM Immunoglobulin domain; Immunoglobulin C region; Alternative splicing;
 KM 3D-structure; Repeat.
 FT NON TER 1 1
 FT DOMAIN 6 98 IG-LIKE 1.
 FT DOMAIN 126 225 IG-LIKE 2.
 FT DOMAIN 234 330 IG-LIKE 3.
 SQ SEQUENCE 335 AA; 36596 MW; FA3382792CBB13C6 CRC64;

Query Match 23.1%; Score 788; DB 1; Length 335;
 Best Local Similarity 45.7%; Pred. No. 3.1e-43;
 Matches 158; Conservative 53; Mismatches 85; Indels 50; Gaps 5;

QY 285 PQALPYAGSGNLTALAE-----KTSKLHDEVNLVVMRAITQLOKNLTCEVWGFTSPKLM 339
 DB 34 PEPTTLTWNSSGLSSGVHTFPALLQSLYTLSSSVTSTWTPQITTCVVAHPAS---- 89
 QY 340 LSLKLENKAKVSKREKPVWVNLNPEAGMCCLSDSQVLLSEIKYKLPMTSTVPCPAP 399
 DB 90 -----STKVDKTIIP-----RVPTQN-----PCP-- 109
 QY 400 EPKSCDKTHTCPELLGGPSVFLFPPPKPDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVD 459
 DB 110 -PHQRVPCAPADLGGPSVFIFFPKTKDVLMLISTSMVTCVVVDSEDDPDVQISMFVN 168
 QY 460 GVEVNAKTKRREQDYNSTYRVSVVLTVLHODMNGEKYCKVKSNKALPAPIEKTISKAK 519
 DB 169 NVEVHTAQQTGTHREDVNSTLRVVSALPIQHDWMGSKFKCKVNNRALPSPIEKTISKPR 228
 QY 520 GQRPPEPVYTLPPSRDELTKNQVSLTCLVNGFYPSDIAVEMESNGOPENNYKTPPVLD 579

```

DB      229 GPRAPQVYVLPPEAEMTKKESLTCMTGFLPAIADWMTNSNGTEQNYKTAIVLDS 288
QY      580 DGSFPLYSKLTVDKSRMQGVNFCSCVMHEALHNHYTKSLSPG 625
DB      289 DGSFYMSKLRVQKSTWERSLFPACSVHEVLHNHLLTTKTSRSLG 334

RESULT 33
GCB_MOUSE
ID      GCB_MOUSE      STANDARD;      PRT;      336 AA.
AC      P01866;
DT      21-JUL-1986 (Rel. 01, Created)
DT      21-JUL-1986 (Rel. 01, Last sequence update)
DT      15-MAR-2004 (Rel. 43, Last annotation update)
DE      Ig gamma-2b chain C region secreted form.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      (1)
RX      MEDLINE=80120716; PubMed=6766534;
RA      Yamawaki-Kataoka Y., Kataoka T., Takahashi N., Obata M., Honjo T.;
RT      "Complete nucleotide sequence of immunoglobulin gamma2b chain gene
RT      cloned from newborn mouse DNA."
RN      Nature 283:786-789(1980).
[2]
RX      MEDLINE=80081501; PubMed=117548;
RA      Tucker P.W., Marcu K.B., Slichter J.L., Blattner F.R.;
RT      "Structure of the constant and 3' untranslated regions of the murine
RT      gamma 2b heavy chain messenger RNA."
RN      Science 206:1299-1303(1979).
[3]
RX      MEDLINE=80081502; PubMed=117549;
RA      Tucker P.W., Marcu K.B., Newell N., Richards J., Blattner F.R.;
RT      "Sequence of the cloned gene for the constant region of murine gamma
RT      2b immunoglobulin heavy chain."
RN      Science 206:1303-1306(1979).
[4]
RX      MEDLINE=82173203; PubMed=6801173;
RA      O'Lo R., Rougeon F.;
RT      "Mouse immunoglobulin allotypes: post-duplication divergence of gamma
RT      2a and gamma 2b chain genes."
RN      Nature 296:761-763(1982).
[5]
RX      CARBOHYDRATE-LINKAGE SITE THR-105.
RX      MEDLINE=94216359; PubMed=7512967;
RA      Kim H., Yamaguchi Y., Maeda K., Matsunaga C., Yamamoto K.,
RA      Irimura T., Takahashi N., Kato K., Arata Y.;
RT      "O-glycosylation in hinge region of mouse immunoglobulin G2b."
RN      J. Biol. Chem. 269:12345-12350(1994).
CC      -1- SUBCELLULAR LOCATION: Secreted (Potential).
CC      -1- ALTERNATIVE PRODUCTS:
CC      Event=Alternative splicing; Named isoforms=2;
CC      Name=Secreted;
CC      IsoId=P01866-1; Sequence=Displayed;
CC      Note=May be the major isoform;
CC      Name=Membrane-bound;
CC      IsoId=P01867-1; Sequence=External;
CC      -1- PTM: O-LINKED GLYCAN CONSISTS OF GAL-GALNAC DISACCHARIDE WHICH IS
CC      MODIFIED WITH 2 SIALIC ACID RESIDUES.
CC      -1- PTM: DISULFIDE BONDS BY HOMOLOGY WITH OTHER IG GAMMA CHAINS.
CC      -1- MISCELLANEOUS: The a allele sequence is shown.
CC      -1- SIMILARITY: Contains 3 immunoglobulin-like domains.
DR      HSSP; P01842; 7FAB.
DR      InterPro; IPR007110; Ig-like.
DR      InterPro; IPR003597; Ig cl.
DR      InterPro; IPR003006; Ig_MHC.

```

```

DR      Pfam; PF00047; Ig; 3.
DR      SMART; SM00407; IGc1; 2.
DR      PROSITE; PSS0835; IG_Like; 3.
DR      PROSITE; PS00290; IG_MHC; 1.
KW      Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
KW      Alternative splicing; Repeat.
FT      DOMAIN 1 6 98
FT      DOMAIN 127 226
FT      DOMAIN 235 331
FT      DISULFID 15 15
FT      DISULFID 27 82
FT      DISULFID 109 109
FT      DISULFID 112 112
FT      DISULFID 115 115
FT      DISULFID 118 118
FT      DISULFID 150 210
FT      DISULFID 256 314
FT      CARBOHYD 105 105
FT      MOD RES 336 336
FT      VARIANT 163 163
FT      VARIANT 194 194
FT      VARIANT 300 300
FT      VARIANT 301 301
FT      CONFLICT 25 25
FT      CONFLICT 36 36
FT      CONFLICT 239 239
SQ      SEQUENCE 336 AA; 36658 MW; 7D879662607C356E CRC64;

Query Match 22.9%; Score 783.5; DB 1; Length 336;
Best local similarity 45.7%; Pred. No. 6; Ie-43;
Matches 159; Conservative 52; Mismatches 84; Indels 53; Gaps 6;

O-LINKED (GALNAC...);
REMOVED POST-TRANSLATIONALLY (PROBABLE).
Q -> R (IN ALLELE B).
T -> A (IN ALLELE B).
N -> D (IN ALLELE B).
M -> I (IN ALLELE B).
L -> S (IN REF. 2 AND 3).
S -> P (IN REF. 2 AND 3).
I -> T (IN REF. 2 AND 3).

285 PQLPQVAGSGNLTALFA-----KTGKLDQEVNLVWMPATOLQKNLTCEWGPISPKLM 339
DB      34 PESTVTVNMGSLSSSVHTFPALLOGLYTWSSSVPSSTWPSQVTCVSHAPAS----- 89
QY      340 LSIKLENKAKYKSKREKPVWNLNPEAGMWCCLSDGQVLLSINIKVLTWSTPPVCPAP 399
DB      90 -----STVDKLEP-----SGPI-----STINCP-- 110
QY      400 EPKSCDKHTC--PELLGSPSVFLPPPKPKDTLMTSRTEPVTCVVVDSHEPEVKNFY 457
DB      111 ---PCKECHKCAPNLEGSPSVIFPPNKKDVLMLSTFKTCTVVVDVEDDDVDQISMF 167
QY      458 VDGVEYHNAKTPREEQVNSTIRVVSVTLVLDQWLNKGKCYKCKVSNKALPAPIETISK 517
DB      168 VNNVEVHTAQOTHRDYNSTIRVSVTLPIQHDWMSGKEPKCKVNNKDLPSPIERTISK 227
QY      518 AKGQREPVYTLPRPRDELTKNOVSLTGLVGVGYSDIAVENESNGQPENNYKTPPVY 577
DB      228 IKGLVAPQVYVLPPEAQLSKRDVSLTGLVGFNPGDLSVENTSGHTEENYKOTAPVL 287
QY      578 DSDGSPFLYSKLTVDKSRMQGVNFCSCVMHEALHNHYTKSLSPG 625
DB      288 DSDGSPFLYSKLMKMTSKWEKTDSPFCNVRHGLKNYLYKKTISRSLG 335

RESULT 34
EPC_MOUSE
ID      EPC_MOUSE      STANDARD;      PRT;      421 AA.
AC      P06336; P01856;
DT      21-JUL-1986 (Rel. 01, Created)
DT      01-FEB-1996 (Rel. 33, Last sequence update)
DT      15-JUL-1999 (Rel. 38, Last annotation update)
DE      Ig epsilon chain C region.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      (1)
RX      MEDLINE=84236092; PubMed=6329728;

```

RA Ishida N., Ueda S., Hayashida H., Miyata T., Honjo T.;
 RT "The nucleotide sequence of the mouse immunoglobulin epsilon gene:
 comparison with the human epsilon gene sequence.";
 RL EMBO J. 1:1117-1123(1982).
 RN [2]
 RP REVISIONS.
 RA Honjo T.;
 RL Submitted (Apr-1986) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 34-421 FROM N.A.
 RX MEDLINE=83117774; PubMed=6618553;
 RA Liu F.-T., Albrandt K., Sutcliffe J.G., Katz D.H.;
 RT "Cloning and nucleotide sequence of mouse immunoglobulin epsilon
 chain cDNA.";
 RL Proc. Natl. Acad. Sci. U.S.A. 79:7852-7856(1982).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X01857; CAA25977.1; .
 DR EMBL; X01857; CAA25978.1; .
 DR PIR; A02144; EHMS.
 DR PIR; A02145; EHMS.
 DR HSSP; P01854; 1IGB.
 DR InterPro; IPR007110; Ig-1-like.
 DR InterPro; IPR003597; Ig_C1.
 DR InterPro; IPR003006; Ig_MHC.
 DR Pfam; PF00047; Ig_4.
 DR SMART; SM00407; IgC1.2.
 DR PROSITE; PS00835; IG_LIKE; 4.
 DR PROSITE; PS00290; IG_MHC; 3.
 DR Immunoglobulin domain; Immunoglobulin C region; Glycoprotein.
 FT NON TER 1 1
 FT DOMAIN 1 90
 FT DOMAIN 91 197
 FT DOMAIN 198 304
 FT DOMAIN 305 421
 FT DISULFID 23 75
 FT DISULFID 121 180
 FT DISULFID 226 285
 FT DISULFID 330 392
 FT CARBOHYD 43 43
 FT CARBOHYD 72 72
 FT CARBOHYD 84 84
 FT CARBOHYD 95 95
 FT CARBOHYD 166 166
 FT CARBOHYD 238 238
 FT CARBOHYD 261 261
 FT CARBOHYD 365 365
 FT CARBOHYD 415 415
 SEQUENCE 421 AA; 47320 MW; 8F909E1F30A06B47 CRC64;

Query Match 10.8%; Score 368.5; DB 1; Length 421;
 Best Local Similarity 27.3%; Pred. No. 2,2e-16;
 Matches 121; Conservative 72; Mismatches 152; Indels 99; Gaps 21;

223 FPLAFTVEKLTGSGGLMWAQERASSKSWITFDLKNKE--VSVKRVTDPPKQMKKPL 280
 DB FPNPVTVA-----WYSD--SLNMSVNEPALGSELKVTSTVSGK--SAKNFTC 75

281 HLTLEQALPOYAGSGNLTFL---ALEAKTGKLGQEVNVLVWRATOLQKLTCEVWGPTSP 336
 DB HVTNR---PSFNESTIILVRPNITREPLELHSSCDNARHST---IQYCFITGHILN 129

337 KLMLSLKLENKE-----AVSKREKRVWVLNPEAGWQ-----CLLSDSGVLT 379
 DB DVSVSMLMDREITDTLQAVTLIKEKGKLASTGCTSKLTITEQOMNSESTFTCKVTSQGVY 189

QY 380 LESNIKVLPTWSTFPCPAPRPSKCDTHNCPBLLGSPSVFLPFPKPKDTLMISRTPEVT 439
 DB LAHTRR-----CPDHEPR-----GVITLIPSPLD-LYQNGAPRLT 225

QY 440 CVVVDV-SHEDPEYKFN-----WYDGVVHNNAKTKPREBOINSTYRVSVLT 487
 DB CLVVDLSEKRVNTWNOEKRTSVASQWY---TKHN-----NATTSITSLPV 272

QY 488 LHODMLNGEKYKCVSNKALPAPIEKTISKAKGPREFQVYTLPSRDELTKNGVSLTCL 547
 DB 273 VAKDWIBSGYQCLVDHDFPKPIVRSITTKPGRSAPVYVFPPEEE-SEDKRTLTCL 331

QY 548 VKGFYPSDIAVEMWNSQPENN--YKTPPYLDSGS---FFLYSKLTVDKSRMQQGVNF 602
 DB 332 IQNFPEDISQWLDGDKLISNSQSHITTP-LKNSGNSQGFIRLEVAKTITWTKRQF 390

QY 603 SCSVMEALHN-HYTKSLSLSPG 625
 DB 391 TCQVTHEALQKPRKLEKTIISTSLG 414

RESULT 35
 EPC_RAT STANDARD; PRT; 429 AA.
 AC P01855;
 DT 21-JUN-1986 (Rel. 01, Created)
 DT 21-JUN-1986 (Rel. 01, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Ig epsilon chain C region.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A. (IMMUNOCYTOMA IR2).
 RC STRAIN=LOU/C/MSL;
 RX MEDLINE=81064537; PubMed=6292865;
 RA Hellman L., Petersen U., Engstrom A., Karlsson T., Bennich H.;
 RT "Structure and evolution of the heavy chain from rat immunoglobulin
 E.";
 RL Nucleic Acids Res. 10:6041-6049(1982).
 RN [2]
 RP SEQUENCE OF 168-342 FROM N.A. (MYELOMA IR162).
 RX MEDLINE=8182019; PubMed=6820340.
 RA Kindsvogel W.R., Reddy E.P., Moore J.M., Faust C.H. Jr.;
 RT "A cloned cDNA probe for rat immunoglobulin epsilon heavy chain:
 construction, identification, and DNA sequence.";
 RL DNA 1:335-343(1982).
 RN [3]
 RP SEQUENCE OF 205-306 FROM N.A.
 RX MEDLINE=82174576; PubMed=6803238;
 RA Hellman L., Petersen U., Bennich H.;
 RT "Characterization and molecular cloning of the mRNA for the heavy
 (epsilon) chain of rat immunoglobulin E.";
 RL Proc. Natl. Acad. Sci. U.S.A. 79:1264-1268(1982).
 CC -1- SIMILARITY: Contains 4 immunoglobulin-like domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; J00744; AAA41379.1; ALT_INIT.
 DR PIR; A93442; EHRT.
 DR HSSP; P01854; 1IGB.
 DR InterPro; IPR007110; Ig-1-like.
 DR InterPro; IPR003597; Ig_C1.
 DR InterPro; IPR003006; Ig_MHC.
 DR Pfam; PF00047; Ig_4.
 DR SMART; SM00407; IgC1.1.

DR PROSITE; PS50835; IG_LIKE; 4.
 KW Immunoglobulin domain; Immunoglobulin C region; Repeat.
 FT NON TER 1 1
 FT DOMAIN 1 89 IG-LIKE 1.
 FT DOMAIN 2 103 IG-LIKE 2.
 FT DOMAIN 3 205 IG-LIKE 3.
 FT DOMAIN 4 314 IG-LIKE 4.
 FT DOMAIN 5 168 R -> N (IN REF. 2).
 FT CONFLICT 168 P -> L (IN REF. 2).
 FT CONFLICT 308
 SQ SEQUENCE 429 AA; 48671 MW; D2970B34EFA72B0 CRC64;

Query Match 10.6%; Score 363.5; DB 1; Length 429;
 Best Local Similarity 30.6%; Pred. No. 4.7e-16;
 Matches 91; Conservative 60; Mismatches 99; Indels 47; Gaps 12;

QY 342 LKLENEAKVSKREKPEVWNPAGMWQ-----CLLSDSGVLLSENIKVLPTWSTPVP 395
 DB 156 IKEGKLASTYSR-----LNIQQQWMSSTFTCKVTSQGE-----NYMAHTR 199
 QY 396 CPAPEPKSCDKYTCPELLGSPVFLPPPKKDTLMISRTPEVTCVVVDVSHEDPEVKEN 455
 DB 200 CSDDPE-----GVITYLIPSPPLD-LYENGTPKLTCLVLDSEEE-NITVT 244
 QY 456 WYVDGVEVNAKTKPREQVNSTYRVSVLTVLHODMLNGEKVKCKVSKNKALEPAIEKTI 515
 DB 245 WVERKKSGISGASORSTKHNNATSTISILPVADKMTBEGGYQCRDHPHFKPIVRSI 304
 QY 516 SKAKGPREQVYT-LPPSRDELTKQVSLTCLVKGFPSYDAVEM--ESNGQPENNYKT 572
 DB 305 TKAPGRSAPEVYVFLPREEE--KDKRTLTCLIQNFPPDISVQWQDSKLPKSGHST 362
 QY 573 TPPLVDSGS---FLYSKLTVDKSRWQGVNFSVMEHALIN-HYTKSLSLSPG 625
 DB 363 TTB-LKYNQSNORFFISRLVETVYKALMTQKQFVHIALREPKLERTISKSLG 418

RESULT 36
 MUCM_MOUSE STANDARD; PRT; 476 AA.
 AC P01873;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Ig mu chain C region membrane-bound form.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NC NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE OF 433-476 FROM N.A.
 RX MEDLINE=80222874; PubMed=6771020;
 RA Early P., Rogers J., Davis M., Calame K., Bond M., Wall R.,
 Hood L.;
 RT "Two mRNAs can be produced from a single immunoglobulin mu gene by
 RT alternative RNA processing pathways.";
 RL Cell 20:313-319(1980).
 RN [2]
 RP SEQUENCE OF 410-476 FROM N.A. (MYELOMA MOPC 104E).
 RX MEDLINE=80222873; PubMed=6771019;
 RA Rogers J., Early P., Carter C., Calame K., Bond M., Hood L.,
 Wall R.;
 RT "Two mRNAs with different 3' ends encode membrane-bound and secreted
 RT forms of immunoglobulin mu chain.";
 RL Cell 20:303-312(1980).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=alternative splicing; Named isoforms=2;
 CC Comment=During differentiation, B lymphocytes switch from
 CC expression of isoform Membrane-bound to isoform Secreted;
 CC Name=Membrane-bound;
 CC IsoId=P01873-1; Sequence=Displayed;
 CC Name=Secreted;
 CC IsoId=P01872-1; Sequence=External;

CC -1- MISCELLANEOUS: The sequence of residues 1-409 is assumed to be
 CC identical with the corresponding region of the secreted form.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; V00821; CAA24202.1; -.
 DR PIR; A02167; MHMSW.
 DR HSSP; P01857; IFCU.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003597; Ig_c1.
 DR InterPro; IPR003006; Ig_MHC.
 DR Pfam; PF00047; Ig; 4.
 DR SMART; SM00407; IgC1; 2.
 DR PROSITE; PS50835; IG_LIKE; 4.
 DR PROSITE; PS00290; IG_MHC; 3.
 KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
 KW Alternative splicing; Transmembrane.
 FT NON TER 1 1
 FT DOMAIN 1 105 CH1.
 FT DOMAIN 2 217 CH2.
 FT DOMAIN 3 218 CH3.
 FT DOMAIN 4 325 CH4.
 FT TRANSMEM 456 473 POTENTIAL.
 FT DISULFID 14 14 INTERCHAIN (WITH LIGHT CHAIN) (PROBABLE).
 FT DISULFID 28 89 BY SIMILARITY.
 FT DISULFID 136 199 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
 FT DISULFID 216 216 BY SIMILARITY.
 FT DISULFID 246 305 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
 FT DISULFID 293 293 BY SIMILARITY.
 FT DISULFID 353 415 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
 FT CARBOHYD 46 46 N-LINKED (GLCNAC. . .).
 FT CARBOHYD 211 211 N-LINKED (GLCNAC. . .).
 FT CARBOHYD 243 243 N-LINKED (GLCNAC. . .).
 FT CARBOHYD 281 281 N-LINKED (GLCNAC. . .).
 SQ SEQUENCE 476 AA; 52656 MW; 8D476575A5204071 CRC64;

Query Match 10.5%; Score 359.5; DB 1; Length 476;
 Best Local Similarity 22.7%; Pred. No. 9.7e-16;
 Matches 127; Conservative 80; Mismatches 187; Indels 165; Gaps 19;

QY 144 ESPPGSSPVQCRSP-RGNIOGKTLVSQELQDSGTWTCTVLQNKQKVEFKDVIYL 202
 DB 3 QSEPNVFLPVSCSPSLSDKLVAMGLAPLPSTISFTWN----- 43
 QY 203 AFQKASSIVYKKEGEQVEFSPLATFVEKLTSGGELMQAERASSSKMTFPLKNEYS 262
 DB 44 -----YONNTEVIO-----GIRTEPPLTGTGKYLA 68
 QY 263 VKRVTPDKLQW-----GKKLPLHLTPQALPYAGSGNLTALAEAKTKYL 308
 DB 69 TSCVLLSPKSLIEGSEYLVCKIHGGKRDHLVPLP-----AAVEM 110
 QY 309 HOEVLNV-----MRATLOKRLTCEVWGPTSPKMLSLKLENKAKVSKREPVWLN 362
 DB 111 NPVVNVFVPPROFSGPAPRKSRLCEATNFFPKPIVSWLMDKGLVSEGFDDPYIEN 170
 QY 363 ----PFA-----GMQCLSDSGVLLSENIKVLPTWSTPVCAPPEP 401
 DB 171 KGSPTQTYKVISLTLTISEIDWNLNLYTCRVHRRGLTFLK----- 210
 QY 402 KSGDKHTGCELLGGS-----VLPPEPKKDTLMISRTPEVTCVVVDVSHEDPEVKMY 457
 DB 211 ---WVSGTC---AASPTDILFTTIPSPAD-IFLSKANLTLVSNLAYE-TLINSWA 262
 QY 458 VDGVEVNAKTKPREQVNSTYRVSVLTVLHODMLNGEKVKCKVSKNKALEPAIEKTI 517
 DB 263 SQSGEFLFKIKIMESHNGTSAKGVASVCVEDMNRKFEVCTVTHRDLPSPQKPKISK 302

```

QY 518 AKGPRE-----POVYTLPSRDEL-T-KNOVSLTCLVKGFPYSDIAVWESNGQ--PENN 569
D 323 ----PNEVHKHPAVYLLPAPAREQNLNRESATVTCIVGFSFPAISVQWLGROGLDPEK 378
QY 570 YKTPPVLD--SDGSFLLYSKLTVDKSNWQGNVPSGCVMEALHNHTQKSLSPQLQ 627
D 379 YVTSAPMEPCAPGPGYFTHSILTVTEEMNSGETYTCVAGHEALPHLVTERTVDS---- 434
QY 628 LDETCABQDGLDGLWTT 646
D 435 -TEGEVNAEERGFEENLWTT 452

RESULT 37
MUC_HUMAN STANDARD; PRT; 454 AA.
ID MUC_HUMAN STANDARD; PRT; 454 AA.
AC P01871;
DT 21-JUN-1986 (Rel. 01, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ig mu chain C region.
GN IGM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
RN (1)
RP SEQUENCE OF 1-434 FROM N.A.
RX MEDLINE=90332450; PubMed=2115996;
RA Friedlander R.M., Nusseznweig M.C., Leder P.;
RT "Complete nucleotide sequence of the membrane form of the human IgM
RT heavy chain.", Res. 18:4278-4278 (1990).
RL Nucleic Acids Res. 18:4278-4278 (1990).
RN (2)
RP SEQUENCE (WALDENSTROM'S MACROGLOBULIN GAL).
RX MEDLINE=75059123; PubMed=4803843;
RA Waranabe S., Barnikol H.U., Horn J., Berttram J., Hilschmann N.;
RT "The primary structure of a monoclonal IgM-immunoglobulin
RT (macroglobulin Gal.), II: the amino acid sequence of the H-chain (mu-
RT type), subgroup H III. Architecture of the complete IgM-molecule.",
RL Hoppe-Seyler's Z. Physiol. Chem. 354:1505-1509 (1973).
RN (3)
RP REVISIONS (GAL).
RX MEDLINE=81066716; PubMed=6777162;
RA Milnesco E., Barnikol-Watanabe S., Barnikol H.U., Milnesco C.,
RA Hilschmann N.;
RT "The primary structure of the constant part of mu-chain-disease
RT protein BOT.", Eur. J. Biochem. 111:275-286 (1980).
RN (4)
RP SEQUENCE (WALDENSTROM'S OU), DISULFIDE BONDS, AND CARBOHYDRATES.
RX MEDLINE=74005511; PubMed=4742735;
RA Putnam F.W., Florent G., Paul C., Shindou T., Shintzu A.;
RT "Complete amino acid sequence of the mu heavy chain of a human IgM
RT immunoglobulin.", Science 182:287-291 (1973).
RN (5)
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=82059479; PubMed=6795593;
RA Rabbits T.H., Forster A., Milstein C.P.;
RT "Human immunoglobulin heavy chain genes: evolutionary comparisons of
RT C mu, C delta and C gamma genes and associated switch sequences.",
RL Nucleic Acids Res. 9:4509-4524 (1981).
RN (6)
RP SEQUENCE OF 299-387 AND 438-454 FROM N.A.
RX MEDLINE=81073306; PubMed=6777778;
RA Dolby T.W., Devono J., Croce C.W.;
RT "Cloning and partial nucleotide sequence of human immunoglobulin mu
RT chain cDNA from B cells and mouse-human hybridomas.",
CC Proc. Natl. Acad. Sci. U.S.A. 77:6027-6031 (1980).
CC -1- MISCELLANEOUS: All 4 combinations of the S/G and V/G polymorphisms
CC at positions 192 and 216 have been observed in human mu chains.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.ch).
CC -----
DR EMBL; X17115; CA34971.1; ALT SEQ.
DR EMBL; X57086; -? NOT_ANNOTATED_CDS.
DR HSSP; P01857; IFC1.
DR Genew; HGNC:5541; IGHM.
DR MIM; 147020; -.
DR GLYcositedB; P01871; -.
DR GO; GO:0005624; C:membrane fraction; NAS.
DR GO; GO:0003823; F:antigen binding; TAS.
DR GO; GO:0006955; P:immune response; NAS.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003597; IG cl.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; Ig; 4.
DR SMART; SM00407; IGH1; 3.
DR PROSITE; PS00835; IG LIKE; 4.
DR PROSITE; PS00290; IG MHC; 3.
KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
KW Polymorphism.
FT NON_TER 1 1
FT DOMAIN 1 105 CH1.
FT DOMAIN 106 218 CH2.
FT DOMAIN 219 324 CH3.
FT DOMAIN 325 454 CH4.
FT DISULFID 14 14 INTERCHAIN (WITH LIGHT CHAIN).
FT DISULFID 28 88 INTERCHAIN (WITH HEAVY CHAIN).
FT DISULFID 135 198 INTERCHAIN (WITH HEAVY CHAIN).
FT DISULFID 215 215 INTERCHAIN (WITH HEAVY CHAIN).
FT DISULFID 245 304 INTERCHAIN (WITH HEAVY CHAIN).
FT DISULFID 292 292 INTERCHAIN (WITH HEAVY CHAIN IN ANOTHER
SUBUNIT).
FT DISULFID 352 414 INTERCHAIN (WITH HEAVY CHAIN).
FT DISULFID 453 453 N-LINKED (GLCNAC. . .).
FT CARBOHYD 46 46 N-LINKED (GLCNAC. . .).
FT CARBOHYD 210 210 N-LINKED (GLCNAC. . .).
FT CARBOHYD 273 273 N-LINKED (GLCNAC. . .).
FT CARBOHYD 280 280 N-LINKED (GLCNAC. . .).
FT CARBOHYD 441 441 N-LINKED (GLCNAC. . .).
FT VARIANT 192 192 S -> G.
FT VARIANT 192 192 /FTid=VAR_003903.
FT VARIANT 216 216 V -> G (in dbSNP:12365).
FT VARIANT 216 216 /FTid=VAR_003904.
SQ SEQUENCE 454 AA; 49556 MW; 21EC72EAD56922E CRC64;
Query Match 10.5%; Score 358.5; DB 1; Length 454;
Best Local Similarity 27.6%; Pred. No. 1.1e-15;
Matches 83; Conservative 62; Mismatches 111; Indels 45; Gaps 10;
QY 332 GPTSPKLMISIKLENKAKVSKREKPVVNLNPEAGMQLSLDSGQVLE-SNIKVLPTW 390
D 173 GPTTYKVTSLTITESP-----WL---SQSMTCVNDHNGLFFQGNASMCVPDQ 219
QY 391 STPYPCAPPEPKSCDKTHTCPDELIGSPSVLPFPKPKDTLMISRTPEVTCVVDVSHDP 450
D 220 DTAL-----RVFAIPPS-PASIFLTSTKTLCTCLVDTLTYD- 254
QY 451 EVKKNWYVDGVEVNAKTKPREQVNSTYVSVLTVLHODMNGKXKCKVSKALPAP 510
D 255 SVTISWTRNGEAVKTHNISHSHNATPSAVGEASICEPDWNSGERTCLVHTDLPSP 314
QY 511 IEKTSIKAKQP-REPQVYTLPSRDEL-T-KNOVSLTCLVKGFPYSDIAVWESNGQP-- 566
D 315 LKQTSNPKGVAAHRPVIYLLPAPAREQNLNRESATITCIVGFSFPAIVGVQMGROGLPS 374

```

```

OY      567  ENNKPTPEPVDD--SDGSFFLYKSTVYKSNQOCNVNVCSCVMHEALHNHTQKSLSP 624
Db      375  PEKVYTSAPMEPOQAPGRYFAHSILTYVSEEMNIGETVYCVAAHEALPNRYTERTVXDST 434
OY      625  G 625
      435  G 435
Db      435  G 435

RESULT 38
EPC_HUMAN
ID      EPC_HUMAN      STANDARD;      PRT;      428 AA.
AC      P01854;
DT      21-JUL-1986 (Rel. 01, Created)
DT      21-JUL-1986 (Rel. 01, Last sequence update)
DE      10-OCT-2003 (Rel. 42, Last annotation update)
DE      Ig epsilon chain c region.
GN      IGHE.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX      NCBI_Taxid=9606;
      [1]
RN      SEQUENCE FROM N.A.
RX      MEDLINE=83168897; PubMed=6300763;
RA      Seno M., Kurckawa T., Ono Y., Onda H., Sasada R., Igarashi K.,
RA      Kiuchi M., Sugino Y., Nishida Y., Honjo T.;
RT      "Molecular cloning and nucleotide sequencing of human immunoglobulin
RT      epsilon chain cDNA.";
RL      Nucleic Acids Res. 11:719-726(1983).
      [2]
RN      SEQUENCE FROM N.A., AND VARIANT LEU-359.
RP      MEDLINE=83001945; PubMed=6288268;
RX      Max E.E., Battey J., Ney R., Kirsch I.R., Leder P.;
RT      "Duplication and deletion in the human immunoglobulin epsilon genes.";
RL      Cell 29:691-699(1982).
      [3]
RN      SEQUENCE FROM N.A.
RP      MEDLINE=84236029; PubMed=6234164;
RX      Flanagan J.G., Rabbits T.H.;
RT      "The sequence of a human immunoglobulin epsilon heavy chain constant
RT      region gene, and evidence for three non-allelic genes.";
RL      EMBO J. 1:655-660(1982).
      [4]
RN      SEQUENCE FROM N.A.
RP      MEDLINE=84207910; PubMed=6327276;
RX      Ueda S., Nakai S., Nishida Y., Hiseajima H., Honjo T.;
RT      "Long terminal repeat-like elements flank a human immunoglobulin
RT      epsilon pseudogene that lacks introns.";
RL      EMBO J. 1:1539-1544(1982).
      [5]
RN      PRELIMINARY SEQUENCE (MYELOMA PROTEIN ND).
RP      Bernich H.H., Johansson S.G.O., von Bahr-Lindstrom H.;
RA      (in) Bach M.K. (eds.);
RL      Immediate hypersensitivity: modern concepts and developments, pp.1-36,
      Marcel Dekker, New York (1978).
      [6]
RN      SEQUENCE OF 1-40; 68-114 AND 427-428 FROM N.A.
RP      MEDLINE=83065234; PubMed=6815556;
RX      Kerten J.H., Molgaard H.V., Houghton M., Derbyshire R.B., Viney J.,
RA      Bell L.O., Gould H.J.;
RT      "Cloning and sequence determination of the gene for the human
RT      immunoglobulin epsilon chain expressed in a myeloma cell line.";
RL      Proc. Natl. Acad. Sci. U.S.A. 79:6661-6665(1982).
      [7]
RN      3D-STRUCTURE MODELING.
RP      MEDLINE=87089848; PubMed=3796618;
RX      Padlan E.A., Davies D.R.;
RT      "A model of the Fc of immunoglobulin E.";
RL      Mol. Immunol. 23:1063-1075(1986).
      -1- SIMILARITY: Contains 4 immunoglobulin-like domains.
      -----
      This SWISS-PROT entry is copyright. It is produced through a collaboration
      CC
  
```

```

CC      between the Swiss Institute of Bioinformatics and the EMBL outstation on its
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.1ab-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
CC      -----
CC      EMBL, L00022; AAB59424.1; ALT_INIT.
DR      PIR; A22771; EHHU.
DR      PDB; 1IGE; 15-JUL-92.
DR      PDB; 1EP5; 30-JAN-02.
DR      PDB; 1G84; 16-MAY-01.
DR      PDB; 1OOV; 18-SEP-02.
DR      Genew; HGNC:5522; IGHE.
DR      MIM; 147180; -.
DR      GO; GO:0003823; P:antigen binding; NAS.
DR      GO; GO:0006955; P:immune response; NAS.
DR      InterPro; IPR007110; Ig-like.
DR      InterPro; IPR003597; Ig-cl.
DR      InterPro; IPR003006; Ig_MHC.
DR      Pfam; PF00047; Ig; 4.
DR      SMART; SM00407; IGc1; 4.
DR      PROSITE; PS00835; IG_LIKE; 4.
DR      PROSITE; PS00290; IG_MHC; 3.
KW      Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
KW      3D-structure; Repeat.
FT      NON TER 1
FT      DOMAIN 6 103 IG-LIKE 1.
FT      DOMAIN 112 210 IG-LIKE 2.
FT      DOMAIN 214 318 IG-LIKE 3.
FT      DOMAIN 324 423 IG-LIKE 4.
FT      DISULFID 14 14 INTERCHAIN (WITH A LIGHT CHAIN).
FT      DISULFID 15 105 INTERCHAIN (WITH A HEAVY CHAIN).
FT      DISULFID 29 85 INTERCHAIN (WITH A HEAVY CHAIN).
FT      DISULFID 121 121 INTERCHAIN (WITH A HEAVY CHAIN).
FT      DISULFID 135 193 INTERCHAIN (WITH A HEAVY CHAIN).
FT      DISULFID 209 299 INTERCHAIN (WITH A HEAVY CHAIN).
FT      DISULFID 239 299 INTERCHAIN (WITH A HEAVY CHAIN).
FT      DISULFID 345 405 INTERCHAIN (WITH A HEAVY CHAIN).
FT      CARBOHYD 21 21 N-LINKED (GLCNAC. . .).
FT      CARBOHYD 49 49 N-LINKED (GLCNAC. . .).
FT      CARBOHYD 99 99 N-LINKED (GLCNAC. . .).
FT      CARBOHYD 146 146 N-LINKED (GLCNAC. . .).
FT      CARBOHYD 252 252 N-LINKED (GLCNAC. . .).
FT      CARBOHYD 275 275 N-LINKED (GLCNAC. . .).
FT      VARIANT 359 359 W -> L (POSSIBLE POLYMORPHISM).
FT      /FtId=VAR_003885.
FT      STRAND 110 110
FT      STRAND 113 117
FT      HELIX 122 124
FT      TURN 125 125
FT      STRAND 130 141
FT      STRAND 146 150
FT      STRAND 155 155
FT      STRAND 159 161
FT      STRAND 165 166
FT      TURN 168 169
FT      STRAND 172 181
FT      HELIX 182 186
FT      TURN 187 188
FT      STRAND 192 196
FT      TURN 198 199
FT      TURN 201 202
FT      STRAND 205 207
FT      STRAND 218 222
FT      TURN 226 229
FT      TURN 231 232
FT      STRAND 236 242
FT      STRAND 252 252
FT      STRAND 255 257
FT      TURN 258 259
FT      STRAND 260 260
FT      TURN 265 266

```

```

FT STRAND 271 274
FT TURN 275 276
FT STRAND 277 285
FT HELIX 288 293
FT TURN 294 294
FT STRAND 297 302
FT TURN 304 305
FT STRAND 310 314
FT STRAND 322 322
FT STRAND 325 329
FT HELIX 334 337
FT STRAND 340 348
FT STRAND 350 351
FT STRAND 356 360
FT TURN 366 369
FT STRAND 374 374
FT STRAND 377 378
FT TURN 380 381
FT STRAND 384 385
FT STRAND 387 393
FT HELIX 394 398
FT TURN 399 400
FT STRAND 404 408
FT STRAND 418 421
SQ SEQUENCE 428 AA; 47019 MW; 25C4CA072AAS58A0 CRC64;

Query Match 10.4%; Score 356; DB 1; Length 428;
Best Local Similarity 26.6%; Pred. No. 1.4e-15;
Matches 118; Conservative 69; Mismatches 154; Indels 102; Gaps 18;

QY 267 TQDPKLT-----OMGKLPPLH-----LPQALPOVAGSGLN-----TLAEKTKGL 308
DB 3 TQSPSVFPLTCCCKNIPSNATSVTLGLATGYFEPVAVVMDTSGLSMTWTLPTATITLT 62
QY 309 ---HOEVNLVVMRATOLQKNLTCEV-WGPTSPKMLSLKENKAKVSKRE---KPVVVL 361
DB 63 SGHVAITSLITVSGAMAKQMTFCVVAHTPSSSTDW-----VDNKTFVSCSRDFTPTVKIL 117
QY 362 N-----PEAGMWOCLLS-----DSGOVLSSNKKVLPWSTPVPKPAPE 400
DB 118 QSSCDGGGHPPTTQLCLTVSGYTPGTINITWLEGGQVMDV-----LSTASTOGBELAS 173
QY 401 PKS-----CDKHTTC-----PELGGPSVFLFPPKPKDITL 430
DB 174 TQSEDLTSLQKMLSDRTTCQVYOGHTFEDSTKKCADSNR---GVSAYLSRSPFD-L 229
QY 431 MISRTPEVTVVVDVSHEDPEVKENYVDGVEVHNAKTKPREEOYNSTYRVSVYLVTLHQ 490
DB 230 FIRKSPITCLVVDLAPSKGTVNLTWRSASGKPVVHSTRKEKQNGTLVTSILPVGTR 289
QY 491 DWLNGKEYCKVSKNALPAPLEKTISSKAGQPREPOVYTL-----PPSRDELTKNOVSLT 545
DB 290 DWIGETVQCRVTHPHLPRALMRSTTKTSGPRAPAEVVAFAATPEWPGSRDRT-----LA 344
QY 546 CLVNGEYPSDIJAVEWESN---GQPENNYKTPPVLDSDSFLYSKLTJYDKSRMOQGNVES 603
DB 345 CLIQNFMEDISVQWLNHEVOLPDKRHSTTTPRKTKSGGFVFSKLEVTRAEMQKDEFI 404
QY 604 CSVNHGALAHNYT-QKSLSLSPG 625
DB 405 CRAVHEAASPSQTVQRAVSVNPG 427

```

```

OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RX NCBI_Taxid=9986;
RN [1]
RP SEQUENCE FROM N.A. (A2 ALLOTYPED).
RX MEDLINE=84088930; PubMed=6418803;
RA Bernstein K.E., Alexander C.B., Reddy E.P., Mage R.G.;
RT "Complete sequence of a cloned cDNA encoding rabbit secreted mu-chain
of Vha2 allotype: comparisons with Vha1 and membrane mu sequences.";
RL J. Immunol. 132:490-495 (1984).
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Comment-During differentiation, B lymphocytes switch from
CC expression of isoform Membrane-bound to isoform Secreted;
CC Name=Membrane-bound;
CC IsoId=P04221-1; Sequence=Displayed;
CC Name=Secreted;
CC IsoId=P03988-1; Sequence=External;
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ch/announce/
CC or send an email to license@ebi.ebi.ch).
CC -----
DR EMBL; K01357; AAA31293.1; -.
DR PIR; A02165; MHRBM.
DR HSSP; P01842; 7FAB.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003597; Ig cl.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig; 4.
DR SMART; SM00407; Igcl; 2.
DR PROSITE; PS50835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 3.
DR Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
KW Alternative splicing; Transmembrane.
FT NON_TER 1 1
FT DOMAIN 1 106 CH1.
FT DOMAIN 107 222 CH2.
FT DOMAIN 223 327 CH3.
FT DOMAIN 328 458 CH4.
FT TRANSMEM 459 476
FT DISULFID 14 14
FT DISULFID 28 90
FT DISULFID 137 200
FT DISULFID 219 219
FT DISULFID 249 308
FT DISULFID 296 296
FT DISULFID 356 418
FT CARBOHYD 46 46
FT CARBOHYD 114 114
FT CARBOHYD 212 212
FT CARBOHYD 261 261
FT CARBOHYD 277 277
FT CARBOHYD 284 284
SQ SEQUENCE 479 AA; 52351 MW; 689C637A47BE19FC CRC64;

Query Match 10.4%; Score 355.5; DB 1; Length 479;
Best Local Similarity 24.8%; Pred. No. 1.6e-15;
Matches 122; Conservative 81; Mismatches 176; Indels 113; Gaps 21;

QY 223 FPL-----AFTEVKTSGGELMWAQERASSKSMITFDLKNKEVSVKRVTDQPKLQMGKK 277
DB 9 YPLVSCGALTGDNLVAMGCLARDPLSSVTFSW-SFK-NSSEISSRTVTFPVPVKRGDK 66
QY 278 -----LPLHLTLPOL-----POYA 292
DB 67 YMATSQVLVPSKDYLTQSTEEYLVCKYQHSNSNBDLRVSPVDSBLPENVSVFIIPRDSFS 126
QY 293 GSGNLTALAEAK-TGKLHOEVNLVVMRATQ-----LQKLTCEVWGPPTSPKMLSLKLE 345

```

```

Db      127 GSGTRKRLICQATGSPKQISVSWLRDQGVESVLTLPKVEATKAGAPATPSISSMLT 186
Qy      346 NKEAKYSKREKPYWVLNPEAGMOCILSDSGVLLSNIKVLTPTWPPCPAPEKSCD 405
Db      187 IRESD-----WL-----SQSLYTCRVDRHG-IFPKKVMSSSEST-----TPSP----- 225
Qy      406 KHTHCCELLGSPVFLPPKPKDTLMISRTPEVTCVVDVSDPEVKFMWYDVGEVHN 465
Db      226 -----GIVVFPIAPSFADT-FLSSARLICLVLTDTTYTG-SLNISSW-----ASHN 268
Qy      466 AKT-----KPREQVNSTYRVSVLTVLHODWLNKGYCKKYNKALPAPIEKTISKAG 520
Db      269 GKALDTHMNTTESHPRNATFSAMGEASVCAEDMESGQFTCTVTHADLPFLKHTISKSH 328
Qy      521 QPRE-POVYTLPPSRBELT-KNOVSLTCLVKGFPSPDIAYEWSNGQP--ENNYKTPPY 576
Db      329 VAKHPAVVYLPAPAREQLVRESATVTCVKGFSPADVFQMOQRQPLSSDCKVTSAA 388
Qy      577 LD--SDGSFELYSLTVDKSRMOQGNVFCSCVMHEALHNHYTQKSLSLSPGLDFTCAE 634
Db      389 PEPQAPGLYFTHTSTLTVEEDMNSGETFTTCVGHFALPHWVTERIVDKS-----TEGEVG 443
Qy      635 AQDELDGLMTT 646
Db      444 AEEGFENLMTT 455

RESULT 40
MUCB_HUMAN          STANDARD;          PRT;          391 AA.
ID   MUCB_HUMAN
AC   P04220;
DT   20-MAR-1987 (Rel. 04, Created)
DI   20-MAR-1987 (Rel. 04, Last sequence update)
DE   16-OCT-2001 (Rel. 40, Last annotation update)
OS   Ig mu heavy chain disease protein (BOT).
OC   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RX   MEDLINE=6418166; PubMed=6425189;
RA   Barnikol-Watanabe S., Mihaesco E., Mihaesco C., Barnikol H.U.,
RA   Hilschmann N.;
RT   "The primary structure of mu-chain-disease protein BOT. Peculiar
RT   amino-acid sequence of the N-terminal 42 positions."
RL   Hoppe-Seyler's Z. Physiol. Chem. 365:105-118(1984).
CC   -1- MISCELLANEOUS: This protein has no V region homology or CH1
CC   region.
DR   PIR; A02163; MHMURT.
DR   HSSP; P01857; 1FC1.
DR   GO; GO:0005624; C:membrane fraction; NAS.
DR   GO; GO:0003823; F:antigen binding; TAS.
DR   GO; GO:0006955; P:immune response; NAS.
DR   InterPro; IPR007110; IG-1like.
DR   InterPro; IPR003597; IG_c1.
DR   InterPro; IPR003066; IG_MHC.
DR   Pfam; PF00047; Ig; 3.
DR   SMART; SM00407; IGC1; 2.
DR   PROSITE; PSS0835; IG_LIKE; 3.
DR   PROSITE; PSS0290; IG_MHC; 3.
KM   Immunoglobulin domain; Immunoglobulin C region.
FT   DOMAIN          1 42     PRE-C-PART (NO V REGION HOMOLOGY).
FT   DOMAIN          43 155    CH2.
FT   DOMAIN          156 261    CH3.
FT   DOMAIN          262 391    CH4.
SQ   SEQUENCE      391 AA; 43057 MW; 9100843AF0CF021A CRC64;

Query Match      10.3%; Score 352.5; DB 1; Length 391;
Best Local Similarity 27.8%; Pred. No. 2,1e-15;
Matches 84; Conservative 61; Mismatches 110; Indels 47; Gaps 11;
Qy      332 GPSPTKMLSLKLENKAKYSKREKPYWVLNPEAGMOCILSDSGVLLSNIKVL--PT 389

```

```

Db      110 GPTTKVSTILIKESD-----WL-----GQSMFTCRVDHRG-LTFQQAASSKCGPD 155
Qy      390 WSPVPCPAPPEKSCDKHTHCCELLGSPVFLPPKPKDTLMISRTPEVTCVVDVSHD 449
Db      156 QDTAI-----RVFAIPPS-FASIFLTSTKTLCLVLTDTTYTD 191
Qy      450 PEVKFMWYDVGEVHNATKPREQVNSTYRVSVLTVLHODWLNKGYCKKYNKALPAPA 509
Db      192 -SVTISWTRQGEAVKTHNISESPNATFSAVGEASICEDDMDGERTCTVTHIDPS 250
Qy      510 PIKTIISKAGQP-REPOVYTLPPSRBELT-KNOVSLTCLVKGFPSPDIAYEWSNGQP- 566
Db      251 PLKQIISRKGVALLRPDYLLPAPAREQLVRESATITLTVGFSRADVFQMOQRQPL 310
Qy      567 -ENNYKTPPYLD--SDGSFELYSLTVDKSRMOQGNVFCSCVMHEALHNHYTQKSLIS 623
Db      311 SPEKYVTSAPWEPQAPGRYFAHSILTVSEEMNNGEYTCVVAHEALPNRVTEDKS 370
Qy      624 PG 625
Db      371 TG 372

RESULT 41
MUC_MOUSE           STANDARD;          PRT;          455 AA.
ID   MUC_MOUSE
AC   P01872;
DT   21-JUL-1986 (Rel. 01, Created)
DI   21-JUL-1986 (Rel. 01, Last sequence update)
DE   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Ig mu chain C region secreted form.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RX   MEDLINE=81076590; PubMed=6255422;
RA   Kawakami T., Takahashi N., Honjo T.;
RT   "Complete nucleotide sequence of mouse immunoglobulin mu gene and
RT   comparison with other immunoglobulin heavy chain genes."
RL   Nucleic Acids Res. 8:3933-3945(1980).
RN   [2]
RX   MEDLINE=82051295; PubMed=6795090;
RA   Goldberg G.I., Vanin E.F., Zrolka A.M., Blattner F.R.;
RT   "Sequence of the gene for the constant region of the mu chain of
RT   Balb/c mouse immunoglobulin."
RL   Gene 15:33-42(1981).
RN   [3]
RX   MEDLINE=81165562; PubMed=6260591;
RA   Aufray C., Rougeon F.;
RT   "Nucleotide sequence of a cloned cDNA corresponding to secreted mu
RT   chain of mouse immunoglobulin."
RL   Gene 12:77-86(1980).
RN   [4]
RX   MEDLINE=79223904; PubMed=111247;
RA   Kehry M.R., Sibley C.H., Fuhrman J.S., Schilling J.W., Hood L.E.;
RT   "Amino acid sequence of a mouse immunoglobulin mu chain."
RL   Proc. Natl. Acad. Sci. U.S.A. 76:2932-2936(1979).
RN   [5]
RX   MEDLINE=83075344; PubMed=6816276;
RA   Kehry M.R., Fuhrman J.S., Schilling J.W., Rogers J., Sibley C.H.,
RA   Hood L.E.;
RT   "Complete amino acid sequence of a mouse mu chain: homology among
RT   heavy chain constant region domains."
RL   Biochemistry 21:5415-5424(1982).
CC   -1- SUBCELLULAR LOCATION: Secreted (Probable).
CC   -1- ALTERNATIVE PRODUCTS:

```

```

CC Event-Alternative splicing; Named isoforms=2;
CC Comment-Editing differential expression, B lymphocytes switch from
CC expression of isoform Membrane-bound to isoform Secreted;
CC Name=Secreted;
CC IsoId=P01872-1; Sequence=Displayed;
CC Name=Membrane-bound;
CC IsoId=P01873-1; Sequence=External;
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; V00827; -; NOT_ANNOTATED_CDS.
DR PIR; A02186; MHMS.
DR HSSP; P01857; 1FC1.
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003597; IG_cl.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; Ig; 4.
DR SMART; SM00407; IgC1; 2.
DR PROSITE; PS50835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 3.
KM Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
KM Alternative splicing.
FT NON_TER 1
FT DOMAIN 1 105 CH1.
FT DOMAIN 106 217 CH2.
FT DOMAIN 218 324 CH3.
FT DOMAIN 325 436 CH4.
FT DOMAIN 437 455 C-TERMINAL SEGMENT OF SECRETED FORM.
FT DISULFID 14 14 INTERCHAIN (WITH LIGHT CHAIN) (PROBABLE).
FT DISULFID 28 89 BY SIMILARITY.
FT DISULFID 136 199 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT DISULFID 216 305 BY SIMILARITY.
FT DISULFID 246 305 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT DISULFID 293 293 BY SIMILARITY.
FT DISULFID 353 415 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT DISULFID 454 454 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT CARBOHYD 46 46 N-LINKED (GLCNAC. . .).
FT CARBOHYD 211 211 N-LINKED (GLCNAC. . .).
FT CARBOHYD 243 243 N-LINKED (GLCNAC. . .).
FT CARBOHYD 281 281 N-LINKED (GLCNAC. . .).
FT CARBOHYD 442 442 N-LINKED (GLCNAC. . .).
FT VARIANT 78 78 S -> N (IN MOPC 104E).
FT VARIANT 101 101 H -> Q (IN MOPC 104E).
FT VARIANT 226 226 T -> N (IN TEPIC183 AND MOPC 104E).
FT VARIANT 258 258 N -> T (IN TEPIC183 AND MOPC 104E).
FT VARIANT 258 258 N -> S (IN TEPIC183).
FT VARIANT 368 368 L -> K (IN TEPIC183 AND MOPC 104E).
SQ SEQUENCE 455 AA; 50101 MM; 4CBES7CB602F9B51 CRC64;
-----
Query Match 10.2%; Score 348; DB 1; Length 455;
Best Local Similarity 22.7%; Pred. No. 4.9e-15;
Matches 122; Conservative 77; Mismatches 179; Indels 160; Gaps 18;
-----
QY 144 ESPPGSSPVQCRSP-RGKNIQGGKTLVSQLELQDSGTWTCTVLQNKVRFIDIVL 202
DB 3 QSPFNVPFLVSCSPSLDKNLVAMGCLARDLPSTISTTN----- 43
QY 203 AFQKASSIVYKKEGQVEFSPLAFVTEKLTGSGELMWQARRASSKSWITFDLKNKEVS 262
DB 44 -----YQNTTEVIO-----GIRTFPLRTGKXLA 68
QY 263 VKRVTQDPKLOM-----GKULPLHLTPQALPQVAGSGNLTALAKTKGKL 308
DB 69 TSQVLLSPKSTLESGDEVLYCKIHVGKGRDLHVPF-----AAVEM 110
QY 309 HOEVNLVV-----MRATQLOKNLTCEVWGPTSPKMLSTLKENKAKVSKREKPVWVLN 362

```

```

DB 111 NPVNVFVPPRDFSGAPRKSCLICEATNFPKPIVSMKQGLVESGETTDPVTIEN 170
QY 363 -----PEA-----GMWCLLSDSGVLLBSNIVKLPWTSPVPCAPAP 401
DB 171 KGSTPQYKYVISTITTEIDMLNLYTCRDYHGLFLPK----- 210
QY 402 KSCDKHTICPELLGSPS-----VFLPPPKDITLMISRTPEVTCVVVDVSHEDPEVKRWY 457
DB 211 ----NVSSTC---AASPSTDLITFTIPSPAD-IFLSASANTLCVSLMAYE-ITLNTSWA 262
QY 458 VDGVEVNAKTKPREEQYNSTYRVSVLTVLDQMLNGKEYKKCKVSKALPAPIEKTKS 517
DB 263 SOSGEPELETIKIKIMESHNGTFSAKGVASCVEDMNRKKEVCTVTRDLPSPQKKFTSK 322
QY 518 AKGQPRE-----POVYTLPPSRDELTKNGVSLTCLVKGFPSPDIAVEMESNGO--PENN 569
DB 323 -----PNEVHKHPPVAVLLPPAREQNLNRESATVTCVKGSPADISVQWLORGQLPQEK 378
QY 570 YKTPPVLD--SDGSFPLYSKLTVDKSRWQGNVFCSVNHEALHNHYTQKSLSLSPG 625
DB 379 YVTSAPMPDEGAPGFYFTTHSLITVTEEMNSGETYTCVGHENALPHLVTERTVDSKSG 436
-----
RESULT 42
MUC_SUNMU STANDARD; PRT; 457 AA.
ID MUC_SUNMU
AC P20768;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ig mu chain C region.
OS Suncus murinus (House shrew) (Musk shrew).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Insectivora; Soricidae; Crociurinae; Suncus.
OX NCBI_Taxid=9378;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver.
RX MEDLINE=69232144; PubMed=2497033.
RA Ishiguro H., Ichihara Y., Namikawa T., Nagatsu T., Kurosawa Y.;
RT "Nucleotide sequence of Suncus murinus immunoglobulin mu gene and
RT comparison with mouse and human mu genes";
RL FEBS Lett. 247:317-322(1989).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X13920; CAA3113.1; ALT_INIT.
DR PIR; S03961; S03961.
DR HSSP; P01842; 7RAB.
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003597; IG_cl.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; Ig; 4.
DR SMART; SM00407; IgC1; 3.
DR PROSITE; PS50835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 3.
KM Immunoglobulin domain; Immunoglobulin C region; Glycoprotein.
FT NON_TER 1
FT DOMAIN 1 105 CH1.
FT DOMAIN 106 220 CH2.
FT DOMAIN 221 326 CH3.
FT DOMAIN 327 457 CH4.
FT DISULFID 13 13 INTERCHAIN (WITH LIGHT CHAIN) (PROBABLE).
FT DISULFID 27 89 BY SIMILARITY.
FT DISULFID 136 200 BY SIMILARITY.
FT DISULFID 217 217 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT DISULFID 248 307 BY SIMILARITY.

```

FT	DISULFID	295	295	INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE)
FT	DISULFID	355	417	BY SIMILARITY.
FT	DISULFID	456	456	INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE)
FT	CARBOHYD	45	45	N-LINKED (GLCNAC . . .) (POTENTIAL)
FT	CARBOHYD	113	113	N-LINKED (GLCNAC . . .) (POTENTIAL)
FT	CARBOHYD	212	212	N-LINKED (GLCNAC . . .) (POTENTIAL)
FT	CARBOHYD	276	276	N-LINKED (GLCNAC . . .) (POTENTIAL)
FT	CARBOHYD	283	283	N-LINKED (GLCNAC . . .) (POTENTIAL)
FT	CARBOHYD	444	444	N-LINKED (GLCNAC . . .) (POTENTIAL)
SC	SEQUENCE	457 AA;	50074 MM;	56C8C086DA4462E9 CR64
Query Match 10.1%; Score 346; DB 1; Length 457;				
Best Local Similarity 24.3%; Pred. No. 6.6e-15;				
Matches 112; Conservative 89; Mismatches 156; Indels 104; Gaps 20;				
QY	204	FOKASSIYKKEGEEVEFSPLAFTVEKLTGSGELMWOAER-ASSSKSWITFDLKN----	258	
DB	43	FNSSSI-----SSQININFEVFGKXMMATSYLLPSTALMGSTDYITCHTGHITGE	97	
QY	259	KEVSKYKVTODPKLQMKGLPLHLTL--POLPYAGSGNLTLEATGKIHQEVNLV	316	
DB	98	KEKV-----ELQVTPDELPPNVSIFVP--PRNSFSGN-----HPTSQLI	135	
QY	317	MRATQKLNCEVW-----GPTSKMLSLKLENKAIVS	352	
DB	136	COASFSFRTIYMSLQGEPEVQSLVSTSAVEAPKSGPTTFVIRLITTE-----	191	
QY	353	KREKVVWLNPEAGMOCLLSDSGVLLSNIKVLPTWSTVPCAPAPKSC--DKITTC	410	
DB	192	-----WLSQRE---FTQALHKG-LTFQKVVSV-----CMGDDTST-	224	
QY	411	PELLGSEVFLFPPPKDTLMSKRPVYTCVVVDVSHEDPEKFWYVDGVEYNAKTP	470	
DB	225	-----GISVFLPLPTFAN-IFLTQSAOLCLVTGATYD-SLIDISWSKONGEALQTHNI	277	
QY	471	REEQYNSRYRVVSVLTVLHODMLKEKCKYSNKLAPLEKITSKKGQPRE--POVYT	529	
DB	278	SESHPNSTFTAKGHASVCREWESEKFTCYQSHDSLSPKQSSRPDYANDPDSVAV	337	
QY	530	LPFSRDEY-TKNQVSLATCLVKGFGYPSDIAVEMESGQP--ENNYKTPPVLD--SDGSFF	584	
DB	338	LPPAEQOLKLRSAISITCLVVDSPDPVFMQGHGQVPDPGHVYTSNPTPEPQRLGYF	397	
QY	565	LYSKLTVDKSRMQGNVFSQSVMEHALNNHYQKLSLSPG	625	
DB	398	VHSILTVSEKDWSSGESFSQVGHALPLPSTVEKAVDXTSG	438	
RESULT 43				
MUC	RABIT	STANDARD;	PRT;	458 AA.
AC	P03988;			
DT	23-OCT-1966 (Rel. 02, Created)			
DT	23-OCT-1966 (Rel. 02, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Ig mu chain C region secreted form.			
OS	Oryctolagus cuniculus (Rabbit).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.			
OX	NCBI_Taxid=9986;			
RN	SEQUENCE FROM N.A. (A2 ALLOTYPED).			
RP	MEDLINE=84088930; PubMed=6418803;			
RX	Bernstein K.E., Alexander C.B., Reddy E.P., Mage R.G.;			
RT	"Complete sequence of a cloned cDNA encoding rabbit secreted mu-chain			
RT	of Vha2 allotype: comparisons with vha1 and membrane mu sequences."			
RL	J. Immunol. 132:490-495(1984).			
CC	-1- SUBCELLULAR LOCATION: Secreted (Probable).			
CC	-1- ALTERNATIVE PRODUCTS:			
CC	Event=Alternative splicing; Named isoforms=2;			
CC	Comment=During differentiation, B lymphocytes switch from			
CC	expression of isoform Membrane-bound to isoform Secreted;			
CC	Name=Secreted;			

```
CC      IsoId=P03988-1; Sequence=Displayed;
CC      Name=Membrane-bound;
CC      IsoId=P04221-1; Sequence=External;
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      The European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; K01357; -; NOT_ANNOTATED_CDS.
DR      PIR; A02164; MHRB.
DR      HSSP; P01842; 7FBA.
DR      InterPro; IPR007110; Ig-like.
DR      InterPro; IPR003597; Ig_c1.
DR      InterPro; IPR003006; Ig_MHC.
DR      Pfam; PF00047; Ig; 4.
DR      SMART; SMO0407; Igcl; 2.
DR      PROSITE; PS50835; Ig_LIKE; 4.
DR      PROSITE; PS00290; IG_MHC; 3.
KW      Immunoglobulin domain; Immunoglobulin C region; glycoprotein;
KW      Alternative splicing.
FT      NON_TER          1
FT      DOMAIN           1    106     CH1.
FT      DOMAIN           107   222     CH2.
FT      DOMAIN           223   327     CH3.
FT      DOMAIN           328   458     CH4.
FT      DISULFD           14    14     INTERCHAIN (WITH LIGHT CHAIN) (PROBABLE).
FT      DISULFD           28    90     BY SIMILARITY.
FT      DISULFD          137   200     BY SIMILARITY.
FT      DISULFD          219   219     INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT      DISULFD          249   308     BY SIMILARITY.
FT      DISULFD          296   296     INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT      DISULFD          356   418     BY SIMILARITY.
FT      CARBOHYD          46    46     N-LINKED (GLCNAC . . ) (POTENTIAL).
FT      CARBOHYD         114   114     N-LINKED (GLCNAC . . ) (POTENTIAL).
FT      CARBOHYD         212   212     N-LINKED (GLCNAC . . ) (POTENTIAL).
FT      CARBOHYD         261   261     N-LINKED (GLCNAC . . ) (POTENTIAL).
FT      CARBOHYD         277   277     N-LINKED (GLCNAC . . ) (POTENTIAL).
FT      CARBOHYD         284   284     N-LINKED (GLCNAC . . ) (POTENTIAL).
FT      CARBOHYD         445   445     N-LINKED (GLCNAC . . ) (POTENTIAL).
SQ      SEQUENCE        458 AA; 49897 MW; F33BD6A3B28E3864 CXC64;
Query Match       10.1%; Score 346; DB 1; Length 458;
Best Local Similarity 24.8%; Pred. No. 6.7e-15;
Matches 117; Conservative 78; Mismatches 168; Indels 108; Gaps 20;
OY      223 FPL-----AFVVEKLTSGGELMWAERSSSKSWITPLKKKEVSVKRVTDQPKLGMRKK 277
DB      9 YPLVSCGALTDGNLVAMVGCIARDFLPSVTFSW-SFK-NNSEISSRIVRTFPVVRKGDK 66
OY      278 -----LPHLLTPQAL-----POYA 292
DB      67 YMATISQVLNPKDYLVQGTEBYLVCKVQHSHNSNRDLRVSPVDSELPPVAVSFIPRRDFS 126
OY      293 GSGNLTALLAEAK-TGKLHDEVNLVVMRAQT-----LOKNLTCEVWGSPTRMLSLKLE 345
DB      127 GSGRKSLIGQAQGFSPKQIJSVEMLRGQGVESGLVTKPEALETKGAGPATFSISMULT 186
OY      346 NKEAKVSKREKRPVNVLPNEAGMQCLSDSQGVLLIESNIKVLPTWSTVPCPAEPKSCD 405
DB      187 ITESD-----WL-----SGLLYTCRDVDHRG-IFFDKNVSMSSSECT-----TPSP---- 225
OY      406 KTHICPELLGGSPVLEFPFRKIDTLMTSRTEVCVVVDVDEHDEPVKFNWVVDGEVYN 465
DB      226 -----GIQVFPLAPSPADTT-FLSKSARLLICLVLDLTYTG-SLNISW-----ASHN 268
OY      466 AKT-----KPREEOYNSTRYRVVSVLTVLVHQOVLNGEKRYCKVSNKALPAPIETISKANG 520
DB      269 GKALDDTHNNITESHNPATFSAFMGAESCAEDMESEGTCTVYTHADDLPPLKHTISKRE 328
```



```

Query Match      10.0%; Score 340; DB 1; Length 450;
Best Local Similarity 25.9%; Pred. No. 1.6e-14;
Matches 100; Conservative 76; Mismatches 142; Indels 68; Gaps 15;

QY 278 LPLHLTLPOAL-----POYAGSGN---LTLALEAKTKGLHQBVLVWMBATOLQKNLT-- 327
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 102 LQVMTLPPREVSFFIPRDAFGBPRKSQULICASGFSFRQWSL--RDGKQIESGVTN 159
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 328 -----CEVWGPTSPKMLSLKLENKAKYVKREKPVWVLNPEAGMOCILSDSGVLLSS 382
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 160 EVZAZAKZSGPTTYKYKTSMLTI-----QEDMWL---SQSVPTCKVEHNG--LTFQO 205
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 383 NIKVLPMTSTPVPCCPAPEPKSCDKHTHCPCELLGSPSVLPFPKPKDTLMISRPEVTCV 442
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 206 NASSMCTISQPV-----GISFTTPPS--PASINTYSAKLSCLV 243
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 443 VDVSHDEPEVKFNWYVDGVEVHNAAKTPREBOYNSTYRVSVLTVLHODMLNKEKXCKV 502
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 244 TDLATYD-SVTLSWTEENGALTKHTNISHSHNGTFSAMGEATVCVEEMESGEQFTCTV 302
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 503 SNKALPAPIEKTISKAKG-QPREPOVYTLPPSRDEL-TKNQVSLTLVKGFPSPDIAYEM 560
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 303 THYDLPVSVKQITISRKGVAVHNPVYVLPSPREQLDLRESATLSCLVGYSPDPVQVM 362
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 561 ESMGO--PENNYKTPPEVLD--SDGSFPLYSKLTVDKSRHQQGNVPSCSVMHEALHNHYT 616
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 363 VQKGQVPPDSVYTSAPMEPPQAPGLYFAHSILTVSEEMNAGETYTCVAHESLPRVT 422
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 617 OKSLSLSPG-----LQIDETCAE 634
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 423 ERSVDKSTGKPTLYNYSVLVSDTAGZ 448
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 46
ALC_RABIT      STANDARD;      PRT;      299 AA.
ID 1-ALC_RABIT
AC P01879;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ig alpha chain C region (Fragment).
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxId=9986;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=84144059; PubMed=6322114;
RA Knight K.L., Martens C.L., Stoklosa C.M., Schneiderman R.D.;
RT "Genes encoding alpha-heavy chains of rabbit IgA: characterization of
   RT cDNA encoding IgA-g subclass alpha-chains.";
RL Nucleic Acids Res. 12:1657-1670(1984).
CC -1- FUNCTION: Ig alpha is the major immunoglobulin class in body
   CC secretions. It may serve both to defend against local infection
   CC and to prevent access of foreign antigens to the general
   CC immunologic system.
CC -1- MISCELLANEOUS: This immunoglobulin belongs to the IgA-G subclass.
   CC It was isolated from a rabbit homozygous FOR A2, N80, DE12, 15,
   CC F71, G75 heavy chain haplotype.
CC -1- SIMILARITY: Contains 2 immunoglobulin-like domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X00353; CAA25100.1; -.
CC PIR; A02174; AHRB.
CC HSSP; P01857; 1FC1.
CC InterPro; IPR007110; Ig-like.

```

```

DR InterPro; IPR003597; Ig_c1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig1.2.
DR SMART; SM00407; Ig1.2.
DR PROSITE; PS00835; IG_LIKE.2.
DR PROSITE; PS00290; IG_MHC.2.
KW Immunoglobulin domain; Immunoglobulin C region; Repeat.
FT DOMAIN 1 167 IG-LIKE 1.
FT DOMAIN 174 276 IG-LIKE 2.
SQ SEQUENCE 299 AA; 32256 MW; 2512FE3E29A223 CRC64;

Query Match      9.0%; Score 306; DB 1; Length 299;
Best Local Similarity 32.2%; Pred. No. 1.3e-12;
Matches 88; Conservative 42; Mismatches 115; Indels 26; Gaps 13;

QY 369 QCLSDSGVLLSESNIKVLPMTSTPVP--CPAPEPKSCDKHTCP-----ELIGG--PSV 419
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 20 QCLGQKSAACHVEYNSVINESLPVPFPDCCPA---NSC---CTCPSSSRNLISGQPSL 73
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 420 FLFPKPKDTLMISRPEVTCVVDVSHDEPEVKFNWYVDGVEVHNAAKTPREBOYNSTY 479
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 74 SLQRPDLGD-LLLGDASLTCTLSGLKPEDAV-FTW--EPTNGNEPVQORQDISGCT 129
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 480 RVSVVLTVLHODMLNKEKCKVSNKALPA-DIEKTISKAKQPREPOVYTLPPSRDEL 538
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 130 SVSVLPSSAEFTWKAETCTVTHTPEIDGSLTATISGVTP--FOVHLPPREBELA 187
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 539 KN-QVSLTCLVKGFPSPDIAYEMESNGQ--PENNY--KTPPEVLDSDGSFPLYSKLTVD 592
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 188 LNEQVTLCTLVGFGSPKDVLSMRHQGEVPPDSFLVWKMSPSSQDKAATVATISLRLVP 247
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 593 KSRMOGNVPSCSVMHEALHNHYTOKSLSLSPG 625
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 248 AEDMNGDTYSCVMGHEGLAEHPTOKTIDRLAG 280
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 47
HVCN_HETFR      STANDARD;      PRT;      461 AA.
ID HVCN_HETFR
AC P23088;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig heavy chain C region, membrane-bound form (Clone 3050).
OS Heterodontus francisci (Horn shark).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Galeomorphii; Heterodontidae; Heterodontiformes;
OC Heterodontidae; Heterodontus.
OX NCBI_TaxId=7792;
RN (1)
RP SEQUENCE FROM N.A.
RX TISSUE=Spleen;
RX MEDLINE=88328985; PubMed=3138109;
RA Kokubu F., Hinde K., Litman R., Shambloot M.J., Litman G.W.;
RT "Complete structure and organization of immunoglobulin heavy chain
   RT constant region genes in a phylogenetically primitive vertebrate.";
RL EMBO J. 7:1979-1988(1988).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X07781; CAA30614.1; ALT_SEQ.
CC PIR; S01854; HYRKO.
CC HSSP; P01842; 7FAB.
CC InterPro; IPR007110; Ig-like.
CC InterPro; IPR003597; Ig_c1.
CC InterPro; IPR003006; Ig_MHC.

```

DR Pfam; PF00047; Ig; 4.
 DR SMART; SM00407; IGL; 2.
 DR PROSITE; PSS0835; IG LIKE; 4.
 DR PROSITE; PSS0290; IG_MHC; 3.
 DR Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
 KW Alternative splicing; Transmembrane.
 FT NON_TER 1 1
 FT DOMAIN 1 99 CH1.
 FT DOMAIN 100 205 CH2.
 FT DOMAIN 206 308 CH3.
 FT DOMAIN 309 418 CH4.
 FT TRANSMEM 438 458
 FT CARBOHYD 164 164 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 245 245 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 275 275 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 374 374 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 411 411 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 415 415 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 437 437 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 461 AA; 50762 MW; 029E91C88D5D911 CRC64;
 Query Match 8.9%; Score 303; DB 1; Length 461;
 Best Local Similarity 25.9%; Pred. No. 3.7e-12;
 Matches 123; Conservative 69; Mismatches 165; Indels 118; Gaps 25;
 QY 208 SSIYVKEGEQVEF---SFPLAFT-----VEKLTGSGELMWQERASSSSKSW 251
 DB 37 TSVSKKDNBPITGLKTYPSVLNKGKTYTQSQTITTESEVSSKITYCEVRGDS--VW 94
 QY 252 ITFDLKNKEVSVKRVTDPLQMGKPLHLTLPLQALPOYAGSNTLALAE----- 303
 DB 95 I-----KEI-----PDCK--GDKVPVILITQSSSEETISRFAVLCSIIDPHES 139
 QY 304 -----KTGKLHGEVNLVWRATQLOKNLTCEVWGSPSPKMLSLKLENKAKVSKREK 357
 DB 140 ITVSWLKDQO-HMESGFVT-----SPTCGVNGTFSAISRIT-----VPARE-- 179
 QY 358 VWVNLPEAGMOCCLSDSGOVLLESNIKVLPTWSTPVPCAPAPRPSCKDHTCPBELLGP 417
 DB 180 -WFTNK---VYTCOVSHQG-VTQSRNIT---GSQVPC-----SCN-----DP 213
 QY 418 SVFLPPPKDPLMISRTPEVTCVVDVSHEDPEVKFNWYDGVENVNAKTPREEQYNS 477
 DB 214 VIKLPSPIEQVL-LEATVILTCTV---VSNAPYGVNWSM---TQEQSLKSRLAVGCE 266
 QY 478 TYRVSUVTVLHQMNLNGEKYKCKVSKALPAPLEKTISSAKGP-REPOV-YTLPSRD 535
 DB 267 ADVISTVNISTQAMLSGAEFYCVVNHQDLPPLRASIHKEVKDLREPSVSIILSPAED 326
 QY 536 ELTKNQVSLTCLVGFPSDIAVEMESNGQPEN--NYKTPPVLDSD-GSFPLYSKLTV 592
 DB 327 VSAKRFSLTCLVGFPSRELIVKRTVNDKSVNPGNNTNTEVMENDSSYFISLSLIA 386
 QY 593 KSRMQGNVFSCSVMEHALNHNHYTKSLSPGLDDETCAEADGELDGMITTD 647
 DB 387 AEEMASGASYSYCVGHEAL-----PLKTIINRTVAKSDDS-DHIWIED 428
 RESULT 48
 HVCS_HETFR STANDARD; PRT; 438 AA.
 AC P23087;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Ig heavy chain C region, secreted form (Clone 3050).
 OS Heterodontus francisci (Horn shark).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
 OC Elasmobranchii; Galeomorphi; Heterodontidae; Heterodontiformes;
 OC Heterodontidae; Heterodontus.
 OX NCBI_TaxID=7792;
 RN [1]

RP SEQUENCE FROM N.A.
 RC TISSUE=Spleen;
 RX MEDLINE=88328985; PubMed=3138109;
 RA Kohabu F., Hinds K., Litman R., Shambloet M.J., Litman G.W.;
 RT "Complete structure and organization of immunoglobulin heavy chain
 constant region genes in a phylogenetically primitive vertebrate";
 RL EMBO J. 7:1979-1988(1988).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>)
 CC or send an email to license@ebi.ac.uk.
 CC -----
 DR EMBL; X07781; CA30613.1; ALT_INIT.
 DR PIR; S01853; HVKCS.
 DR HSSP; P01842; 7PAB.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003597; Ig_c1.
 DR InterPro; IPR003006; Ig_MHC.
 DR Pfam; PF00047; Ig; 4.
 DR SMART; SM00407; IGL; 2.
 DR PROSITE; PSS0835; IG LIKE; 4.
 DR PROSITE; PSS0290; IG_MHC; 3.
 DR Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
 KW Alternative splicing.
 FT NON_TER 1 1
 FT DOMAIN 1 99 CH1.
 FT DOMAIN 100 205 CH2.
 FT DOMAIN 206 308 CH3.
 FT DOMAIN 309 418 CH4.
 FT CARBOHYD 164 164 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 245 245 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 275 275 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 374 374 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 411 411 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 415 415 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 425 425 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 438 AA; 47920 MW; 9BD3F3C3BBD7D15B CRC64;
 Query Match 8.8%; Score 301.5; DB 1; Length 438;
 Best Local Similarity 26.3%; Pred. No. 4.3e-12;
 Matches 119; Conservative 68; Mismatches 161; Indels 105; Gaps 23;
 QY 208 SSIYVKEGEQVEF---SFPLAFT-----VEKLTGSGELMWQERASSSSKSW 251
 DB 37 TSVSKKDNBPITGLKTYPSVLNKGKTYTQSQTITTESEVSSKITYCEVRGDS--VW 94
 QY 252 ITFDLKNKEVSVKRVTDPLQMGKPLHLTLPLQALPOYAGSNTLALAE----- 303
 DB 95 I-----KEI-----PDCK--GDKVPVILITQSSSEETISRFAVLCSIIDPHES 139
 QY 304 -----KTGKLHGEVNLVWRATQLOKNLTCEVWGSPSPKMLSLKLENKAKVSKREK 357
 DB 140 ITVSWLKDQO-HMESGFVT-----SPTCGVNGTFSAISRIT-----VPARE-- 179
 QY 358 VWVNLPEAGMOCCLSDSGOVLLESNIKVLPTWSTPVPCAPAPRPSCKDHTCPBELLGP 417
 DB 180 -WFTNK---VYTCOVSHQG-VTQSRNIT---GSQVPC-----SCN-----DP 213
 QY 418 SVFLPPPKDPLMISRTPEVTCVVDVSHEDPEVKFNWYDGVENVNAKTPREEQYNS 477
 DB 214 VIKLPSPIEQVL-LEATVILTCTV---VSNAPYGVNWSM---TQEQSLKSRLAVGCE 266
 QY 478 TYRVSUVTVLHQMNLNGEKYKCKVSKALPAPLEKTISSAKGP-REPOV-YTLPSRD 535
 DB 267 ADVISTVNISTQAMLSGAEFYCVVNHQDLPPLRASIHKEVKDLREPSVSIILSPAED 326
 QY 536 ELTKNQVSLTCLVGFPSDIAVEMESNGQPEN--NYKTPPVLDSD-GSFPLYSKLTV 592
 DB 327 VSAKRFSLTCLVGFPSRELIVKRTVNDKSVNPGNNTNTEVMENDSSYFISLSLIA 386

```

Db      327 VSAQRFLSLTCLVGRFPFRIYFKWTVNDKSVNPNKYKNTVEVAENDNSYFIYSLSLA 386
QY      593 KSRWQGNVFGSCVMEBALHNHTOKSLSPG 625
Db      387 ABEWASGASYSVCVGHAIPLKLTINRTVKNSSG 419

RESULT 49
MUC_CHICK
ID MUC_CHICK STANDARD; PRT; 446 AA.
AC P01875;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE 19 mu chain C region.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
NCBI_TaxID=9031;
OX [1]
RN SEQUENCE OF 80-446 FROM N.A.
RP MEDLINE=8329221; PubMed=6310496;
RA Dahan A., Reynaud C.-A., Weill J.-C.;
RT "Nucleotide sequence of the constant region of a chicken mu heavy
RT chain immunoglobulin mRNA.";
RL Nucleic Acids Res. 11:5381-5389(1983).
RN [2]
RP SEQUENCE OF 1-79 FROM N.A.
RA Weill J.-C.;
RL Submitted (MAR-1986) to the EMBL/GenBank/DBJ databases.
RU -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X01613; CAA25762.1; -.
DR PIR; A02170; MHCH.
DR HSSP; P01857; 1FC1.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003597; Ig-cl.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig; 3.
DR SMART; SM00407; IgC1; 1.
DR PROSITE; PS50835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 3.
KW Immunoglobulin domain; Immunoglobulin C region.
KW NON_TER
FT DOMAIN 1 105
FT DOMAIN 106 209
FT DOMAIN 210 316
FT DOMAIN 317 427
FT DOMAIN 428 446
FT DISUPEID 16
FT DISUPEID 27
FT DISUPEID 135
FT DISUPEID 237
FT DISUPEID 284
FT DISUPEID 344
FT DISUPEID 446
FT CARBOHYD 45
FT CARBOHYD 46
FT CARBOHYD 73
FT CARBOHYD 130
FT CARBOHYD 198
FT CARBOHYD 382

```

```

FT CARBOHYD 433 433 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 446 AA; 48173 MW; 3CB0CE108949BD17 CRC64;
QY Query Match 8.8%; Score 300; DB 1; Length 446;
Best Local Similarity 27.9%; Pred. No. 5,5e-12;
Matches 116; Conservative 61; Mismatches 179; Indels 60; Gaps 21;

QY 247 SSKSWITPDLNKEVSVKRYVODPKLQMGKLPRLHLLTPQALPOVYSGNLTALAEKNG 306
Db 35 SSIAFTWPSNNSSVSGMDVI--PKVISG-----PKYAVSR--IQNKQSERG 77
QY 307 KLHGEVNLVWMTATLOKALICEWGP--TSPK--LMLSLKENK----- 348
Db 78 KKKQPR---CQAHPRGVSVVNNPGIPFPNGPLPLVTMHPREDFEFGFRNASIL 134
QY 349 AKVSKREKP--VWVLN--PEAGMOCQLSDSGQVLEENIKVLP--WST--PVPCPAP 399
Db 135 CQTRGRRRPTEVTWYNGSPVAATAATTATTVGPEVVAESRISVTESEMDTGATFSCVVEG 194
QY 400 EFKSCDKHTC---PELLGSPVFLFPFKKDTMISRPETVCVVVDVSHEDP--EVKFN 455
Db 195 EKNRNTSKMEGCLPEVVOQDIALRVITPSPFD--IFISKATLTCRVSNVWVNDGLEVSW 252
QY 456 WYVDGVEVNAKTKPREEQYNSTYRVSVLTVLHODMLNGEKYCKVSNKALPAPIEKT 515
Db 253 WKKGKLETALGK-RVLSQNGLYTVDGVAIVCASMDGDDGCVCKVNPDLFPWEKNG 311
QY 516 SKAK-GQPREPOVYTLPSRDELTKNQ-VSLTCLVKGYFSPDSIAVESNGQ--PENNYK 571
Db 312 RRTKASNAPRPVYVYVPPPTTEQLNGQRSLVTCMQGFNPPLFLVWMRNGEPLPOSQSV 371
QY 572 TTPPVVDS--DGSFLYKLTVDKSRWQGNVFGSCVMEBALHNHTOKSLSPG 625
Db 372 TSAPMAENPENSVAYSVLGVGAERWAGNVTCLVGHAIPLQLAQKSVDRASG 427

RESULT 50
HVC2_HETFR
ID HVC2_HETFR STANDARD; PRT; 438 AA.
AC P23085;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig heavy chain C region (Clone 12022) (Fragment).
OS Heterodontus francisci (Horn shark).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Galeomorphi; Heterodontidae; Heterodontiformes;
OC Heterodontidae; Heterodontue.
OC NCBI_TaxID=7792;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=8328985; PubMed=3138109;
RA Kokubu F., Hinds K., Litman R., Shambloet M.J., Litman G.W.;
RT "Complete structure and organization of immunoglobulin heavy chain
RT constant region genes in a phylogenetically primitive vertebrate.";
RL EMBO J. 7:1979-1988(1988).
CC -1- SIMILARITY: Contains 4 immunoglobulin-like domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X07784; CAA30617.1; -.
DR PIR; S00980; HYRKC2.
DR HSSP; P01842; 7FAB.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003597; Ig-cl.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig; 4.

```

DR SMART: SMO0407; IGC1: 2.
 DR PROSITE: PS00835; IG-LIKE; 4.
 DR PROSITE: PS00290; IG-MHC; 3.
 KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein; Repeat.
 FT MON TER 1 1
 FT DOMAIN 5 101 IG-LIKE 1.
 FT DOMAIN 108 202 IG-LIKE 2.
 FT DOMAIN 213 305 IG-LIKE 3.
 FT DOMAIN 315 415 IG-LIKE 4.
 FT CARBOHD 166 166 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHD 245 245 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHD 275 275 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHD 374 374 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHD 411 411 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHD 415 415 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHD 425 425 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SO SEQUENCE 438 AA; 47904 MW; 66E7A9C5BBF7AA7B CRC64;

Query Match 8.7%; Score 237.5; DB 1; Length 438;
 Beest Local Similarity 25.2%; Pred. No. 7.7e-12;
 Matches 113; Conservative 72; Mismatches 167; Indels 97; Gaps 19;

QY SSIYKKKGGEVSEFSEFLAFV-----EKLGGELMMQAEKASSSSK 251
 DB 37 TSIYWKQKEPITTLGKIYPSVLNKKGYTNSQTLTTESEVGSKICYEVRGES--LW 94
 QY 252 ITFPLKKNKESVVKRTYDPPKLOMGKKLPLHLTLPOALPOYAGSGNLTALAEKGTGLKHOE 311
 DB 95 I-----KEI-----LDCK---GDIVEPTVILIGSSSEBELTSRRPATVLSIIDFHPS 139
 QY 312 VNLVVMRATQ---LQKNLTCEVWGPTSPKMLSLKLENKAKYKREKPYVNLNPAAG 366
 DB 140 ITVSMLKQGPWDGFTVSPICEVNGNFSATSRLTVPAGE-----WFSNT--- 184
 QY 367 MMOCILSDSGOVLLESNIKVLPTWSTPVPCCAPAPKESCDKHTCTELLGSGSVFLPPKP 426
 DB 185 VYTCOVAHQ-ETQOSRNTT---GSGVPCS-----IGDPYIKLLPISI 222
 QY 427 KDTLMISRTPEVTCVVDVSHDEPVEKKNMYYDGEVHNAKTKPREEOY-----NSTYRV 481
 DB 223 EQVL-LEATVTLTTC---VSNAFYGVNWS-----TQEEKPLAKSELAVNGPGBDSOV 270
 QY 482 VSVLTVLHQDLNKEKYCKVSNKRLPAPIEKTISKAGQP-REPOV-YTLPPSSDELTK 539
 DB 271 ISTVNIISTQAMLSGAEFYCVVSHQDLPRLPRASIHKEEVKDLRBPFSVLLPPAEDEVSAQ 330
 QY 540 NOVSLTICVKGFPEDIAVEMESNQPEN--NYKTTPYVLDSG--SFLYSLGTLVTKSRW 596
 DB 331 RFLSLTLCVRRGSPREIFIKMTVNDKSVNPGYKNTVEAMENDRNSFTIYSLSLIAEEM 390
 QY 597 QQGVNFGCVNHEALHNHYTKSLSPG 625
 DB 391 ASGASYSYCVGHEALPLKTIINRTVAKSG 419

RESULT 51
 ALC2_HUMAN
 ID ALC2_HUMAN STANDARD; PRT; 340 AA.
 AC P01877;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE IG alpha-2 chain C region.
 GN IGH42
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 FX MEDLINE=84130179; PubMed=6421489;
 RA Flanagan J.G.; Lefranc M.-P.; Rabbits T.H.;

```
RT      "Mechanisms of divergence and convergence of the human immunoglobulin  
RT alpha 1 and alpha 2 constant region gene sequences.";  
RL Cell 36:681-688(1984).  
RN [2]  
RP SEQUENCE (BUT) .  
RA MEDLINE=78137069; PubMed=416441;  
RX Torano A., Putnam F.W.;  
RT "Complete amino acid sequence of the alpha 2 heavy chain of a human  
RT Igalpha immunoglobulin of the Azm (2) allotype.";  
RL Proc. Natl. Acad. Sci. U.S.A. 75:966-969(1978).  
RN [3]  
RP SEQUENCE (MYELOMA PROTEIN IAN) .  
RA MEDLINE=79180140; PubMed=286295;  
RX Tsuzukida Y., Wang C.-C., Putnam F.W.;  
RT "Structure of the Azm(1) allotype of human IgA -- a recombinant  
RT molecule.";  
RL Proc. Natl. Acad. Sci. U.S.A. 76:1104-1106(1979).  
RN [4]  
RP REVIEW.  
RX MEDLINE=91054387; PubMed=2241915;  
RA Kerr M.A.;  
RT "The structure and function of human IgA."  
RL Biochem. J. 271:285-296(1990).  
CC -I- FUNCTION: Ig alpha is the major immunoglobulin class in body  
CC secretions. It may serve both to defend against local infection  
CC and to prevent access of foreign antigens to the general  
CC immunologic system.  
CC -I- SUBUNIT: Monomeric or polymeric.  
CC -I- MISCELLANEOUS: The sequence of the Azm(1) allotype is shown.  
CC -I- SIMILARITY: Contains 3 immunoglobulin-like domains.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb.ch/announce/  
CC or send an email to license@isb.gib.ch).  
-----  
CC EMBL; J00221; AAC59396.1; ALT_INIT.  
DR PIR; B22360; B22360.  
DR HSSP; P01810; PRBJ  
DR Genew; HGNC:5479; IGHA2.  
DR MLN; 147000; -.  
DR GO; GO:0005624; C:membrane fraction; NAS.  
DR GO; GO:0003823; F:antigen binding; TAS.  
DR GO; GO:0006955; P:immune response; NAS.  
DR InterPro; IPRO07110; Ig_1like.  
DR InterPro; IPRO03597; Ig_ct.  
DR InterPro; IPRO03006; Ig_MHC.  
DR Pfam; PF00047; Ig; 3.  
DR SMART; SMO0407; Igcl; 2.  
DR PROSITE; PSS0835; IG_LIKE; 3.  
DR PROSITE; PS00290; IG_MHC; 2.  
DR Immunoglobulin domain; Immunoglobulin C region; Glycoprotein; Repeat.  
FW NON TER 1 1  
FT DOMAIN 6 98 IG-LIKE 1.  
FT DOMAIN 112 207 IG-LIKE 2.  
FT DOMAIN 215 317 IG-LIKE 3.  
FT DISULFID 26 85 PROBABLE.  
FT DISULFID 101 101 INTERCHAIN (WITH LIGHT CHAIN) (PROBABLE).  
FT DISULFID 109 109 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).  
FT DISULFID 110 167 PROBABLE.  
FT DISULFID 134 191 PROBABLE.  
FT DISULFID 169 169 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).  
FT DISULFID 179 179 INTERCHAIN (WITH HEAVY CHAIN IN ANOTHER  
SUBUNIT) (POTENTIAL).  
FT DISULFID 237 300 PROBABLE.  
FT DISULFID 339 339 INTERCHAIN (WITH J CHAIN) (PROBABLE).  
FT CARBOHYD 47 47 N-LINKED (GLCNAC. . ).  
FT CARBOHYD 92 92 N-LINKED (GLCNAC. . ).  
FT CARBOHYD 131 131 N-LINKED (GLCNAC. . ).  
FT CARBOHYD 205 205 N-LINKED (GLCNAC. . ).
```

```

CC CARBOHYD 327 327 N-LINKED (GLCNAC. .).
FT VARIANT 93 93 P -> S (IN A2M(2) ALLOTYPED).
FT VARIANT 102 102 /FTID=VAR 003879.
FT VARIANT 102 102 P -> R (IN A2M(2) ALLOTYPED).
FT VARIANT 279 279 /FTID=VAR 003880.
FT VARIANT 279 279 F -> Y (IN A2M(2) ALLOTYPED).
FT VARIANT 296 296 /FTID=VAR 003881.
FT VARIANT 296 296 D -> E (IN A2M(2) ALLOTYPED).
FT VARIANT 326 326 /FTID=VAR 003882.
FT VARIANT 326 326 V -> I (IN A2M(2) ALLOTYPED).
FT VARIANT 335 335 /FTID=VAR 003883.
FT VARIANT 335 335 V -> A (IN A2M(2) ALLOTYPED).
FT SEQUENCE 340 AA; 36508 MM; 9892270756F3276 CRC64;

Query Match
Best Local Similarity 28.1%; Pred. No. 1e-11;
Matches 99; Conservative 46; Mismatches 140; Indels 67; Gaps 15;

QY 333 PTSPKLM-LSLKENKEAKSKR-----EKPVVNLPEAGM-----WQCLSDSQ 377
DB 3 PTSPKVPPLSDSTPDGNNVACLVQGFPPPLSVTSSEQONTARNPPSPQASD 62
QY 378 VLLESNIKVLPTSTP-----VPCPAPKSCDKHTHCPRLGSP 417
DB 63 LYTTSSQTLTPATQCDGKSVTGHVKNYTPSDVTPPCVPPPPC-----C-----HP 112
QY 418 SVFLPPPKKDTLMISRTPEVTCVVVDVSHEDPEVKFNNYVVDGEVHNAAKTRPREBOYNS 477
DB 113 RLSLHRLPALED-LLLSSEANLCTLTGL-NDASGATFTWPPSSGK--SAVQGPBERDLCG 168
QY 478 TYRNVSVLTLDHDMLNKGEYKCKVSNKALPAPIEKTIISKAKGPREPOVYTLPPSRDEL 537
DB 169 CYSVSSVLPCCAPPMWNGEFTCTAHPLEKTLPLTANITKS-ENTREPEVHLLPPSEEL 227
QY 538 TKKQ-VSLTGLVGFYPSDIAVEMESNGQ--PENNYKTPPVLD-SDG--SFLYGLTY 591
DB 228 ALNELVTLTLANGFSPKDVLVWMLQSGSELPREKILTMASRQEPQGTTFVATSLIKY 287
QY 592 DKSRWQGVNFGSVNHEALHNHYTOKSLSLSPG-----IQLDLTC 632
DB 288 AARDWKKGDTFSCMGVGHLEALPLAFTQKTIDRLAGKPTHVNVSVMAEVDGTC 339

RESULT 52
HVC3 HETFR STANDARD; PRT; 393 AA.
AC P23086;
DT 01-NOV-1991 (Rel. 20, Last Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE 1g heavy chain C region (Clone 6121) (Fragment).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Galeomorphii; Heterodontidae; Heterodontiformes;
OC Heterodontidae; Heterodontus.
OX NCBI_TaxID=7792;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Spine;
RX MEDLINE=68328985; Pubmed=1138109;
RA Kokubu F., Hinde K., Litman R., Shamblott M.J., Litman G.W.;
RT "Complete structure and organization of immunoglobulin heavy chain
constant region genes in a phylogenetically primitive vertebrate.";
RL EMOB J. 7:1979-1988(1988).
CC -1- SIMILARITY: Contains 3 immunoglobulin-like domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)

```

```

CC or send an email to license@isb-sib.ch.
CC -----
CC EMBL: X07782, CA30615.1; -.
CC PIR: S01852; HVRKCL.
CC HSSP: P01842; 7FAB.
CC InterPro: IPR007110; IG-1like.
CC InterPro: IPR003597; IG_C1.
CC InterPro: IPR003006; IG_MHC.
CC Pfam: PF00047; Ig1.
CC SMART: SM00407; IgC1.
CC PROSITE: PS00835; IG_LIKE.
CC PROSITE: PS00290; IG_MHC; 3.
CC Immunoglobulin domain; Immunoglobulin C region; Glycoprotein; Repeat.
FT DOMAIN 1 157 IG-LIKE 1.
FT DOMAIN 168 260 IG-LIKE 2.
FT DOMAIN 270 370 IG-LIKE 3.
FT CARBOHYD 119 119 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 155 155 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 200 200 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 230 230 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 329 329 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 366 366 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 370 370 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 380 380 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 380 380 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 393 AA; 43081 MM; 4E44D076972F18B5 CRC64;

Query Match
Best Local Similarity 27.2%; Pred. No. 2.2e-11;
Matches 118; Conservative 63; Mismatches 162; Indels 91; Gaps 23;

QY 211 VYKKGGEVFPFLPFLATYETKLTGSGELMWQMBASSSSWITFDLKNKEVSKRTQDP 270
DB 13 VANKGTYTQ--SSQLTT-ESEVGSSKTYCEVRGGS--VWL-----KEI-----PDC 56
QY 271 KLQMGKPLPLHTLPQALPOYAGSGLTLALBA-----KTGKLHGVNLV 316
DB 57 K--GDKHPVYILVQSSSEELTSRFATVLCSIIDFHPESTIVSLKQGG-HMEGFPY 112
QY 317 MRATOLQKNLCEWGPSPKMLSLKLENKEAKYSKREPVVNLNPEAGMOCCLSDSG 376
DB 113 -----SPTGVNGTFESATSRLT-----VPARE--WFTNK--VYTCQVSHQG 149
QY 377 QVLESNIKVLPTWSTPVPVPCPAPKSCDKHTHCPRLGSPVFLPPPKDTLMISRP 436
DB 150 -VTQSENIT-----GSQVPC-----SCN-----DPYKLPESIBOVL-LEATV 186
QY 437 EYTCVVVDVSHEDPEVKFNNYVVDGEVHNAAKTRPREQYNSTYRVVSVLTVLHQDMLNKG 496
DB 187 TLNCT---VSNAFYGVNWSM---TQEQKSLKSEINQPEBDDSVISTYNISTQAWLSGA 240
QY 497 EYKCKVSNKALPAPIEKTIISKAKGP-REPQV-YTLPPSRDELTKQVSLTCLVKCFPYS 554
DB 241 EFCVVNHODLPTPLASIHKEEVDKLREPSVILLSPAEDVSAGQFLSLTCLVRFSPR 300
QY 555 DIAVEMESNGQPEN--NYKTTTPVLDSD--GSFLYSLKLVDSRWQGVNFGSVNHEAL 611
DB 301 EIFVKWTINDKSVNPGNYNKTVEAMDNSTFIYLSLIAAEWASGASYSCLVGHAI 360
QY 612 HNNHYTQSLSLSPG 625
DB 361 PLKINRTVYKSSG 374

RESULT 53
ALC1 GORGO STANDARD; PRT; 353 AA.
AC P20758;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 1g alpha-1 chain C region.
GN IGHAI.

```

OS Gorilla gorilla gorilla (Lowland gorilla).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euteria; Primates; Catarrhini; Homidae; Gorilla.
OX NCBI_TaxID=9595;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymph node;
RX MEDLINE=89386006; PubMed=2506527;
RA Kawamura S., Omoto K., Ueda S.;
RT "Nucleotide sequence of the gorilla immunoglobulin alpha 1 gene";
RL Nucleic Acids Res. 17:6732-6732(1989).
CC -1- FUNCTION: Ig alpha is the major immunoglobulin class in body
CC secretions. It may serve both to defend against local infection
CC and to prevent access of foreign antigens to the general
CC immunologic system.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to: license@isb-sib.ch).
CC -----
DR EMBL; X15045; CA33147.1; ALT_INT.
DR HSP; P01810; 2PB.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003597; Ig_c1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig_3.
DR SMART; SM00407; IGc1; 2.
DR PROSITE; PS00835; IG_LIKE; 3.
DR PROSITE; PS00290; IG_MHC; 1.
KM Immunoglobulin domain; Immunoglobulin C region; Glycoprotein; Repeat.
FT NON_TER 1 1
FT DOMAIN 1 98
FT DOMAIN 1 25
FT DOMAIN 1 220
FT DOMAIN 1 228
FT DOMAIN 1 330
FT DISULFID 14 14
FT DISULFID 26 85
FT DISULFID 77 101
FT DISULFID 122 122
FT DISULFID 123 180
FT DISULFID 147 204
FT DISULFID 182 182
FT DISULFID 192 192
FT DISULFID 250 313
FT DISULFID 352 352
FT CARBOHYD 144 144
FT CARBOHYD 340 340
SO SEQUENCE 353 AA; 37755 MW; 48208BDB02AC7514 CRC64;
Query Match 8.3%; Score 282.5; DB 1; Length 353;
Beet Local Similarity 26.5%; Pred. No. 5.2e-11;
Matches 92; Conservative 46; Mismatches 136; Indels 73; Gaps 12;
OY TGKLEHENVLVVMTATOL--KNLTCEVWSPPTSLMLSLKLNKAVSRKRPVWLN 362
DB 60 SGLDTTSSQLTLPPTQCPDGSVYCHNHYNP----- 93
OY 363 PEAGWQCCLSDSGVLLSNIKVLPSTVPCAPAPKSCDKHTCPPELLGSPVFLF 422
DB 94 -----SQGVTVPCRVSPRPSPSTPTPTSPPCCH-----PRLSLH 130
OY 423 PPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKRREQDSTYRV 482
DB 131 RPALED-LLTGSSEANLTCTLTGL--RDAGVFTWTPSSGK--SAVEGPPERDLGCYSVS 186

OY 483 SVLTVLHQMVLNGKREYCKVSNKALPAPIEKTISKAGQPREPOVYTLPSRDELTKNQ- 541
DB 187 SVLGGCAEPNHNHGTFTCTAIVPESKTPPLTTLTKS--GNMRPEVHLLPPSEELANEL 245
OY 542 VSLTGLVKGFPYSPDIAYEWESNGQ--PENNYKTTTPVID-SDG--SFFLYSKLTVDXSRW 596
DB 246 VTLTCLARGFSPKDVLVWMLQSGOELPREKYLTVASRQEPGQTTTFVAVTSILRVAAEDW 305
OY 597 OQGVNFGSVWHEALNHNHYTKSLSRG-----LQDELTC 632
DB 306 KKGVTFSCWGEHALPLAFYQKTIIDLAKGPTHVNVSVMAEVDGTC 352
RESULT 54
ALCL_HUMAN
ID ALCL_HUMAN STANDARD; PRT; 353 AA.
AC P01876;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ig alpha-1 chain C region.
GN IGHA1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euteria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84130179; PubMed=6421489;
RA Flanagan J.G., Lefranc M.-P., Rabbits T.H.;
RT "Mechanisms of divergence and convergence of the human immunoglobulin
RT alpha 1 and alpha 2 constant region gene sequences.";
RL Cell 36:681-688(1984).
RN [2]
RP SEQUENCE (MYELOMA PROTEIN BUR), AND DISULFIDE BONDS.
RX MEDLINE=79151016; PubMed=107164;
RA Putnam F.W., Liu Y.-S.V., Low T.L.K.;
RT "Primary structure of a human IgM immunoglobulin. IV. Streptococcal
RT IgM protease, digestion, Fab and Fc fragments, and the complete
RT amino acid sequence of the alpha 1 heavy chain.";
RL J. Biol. Chem. 254:2865-2874(1979).
RN [3]
RP SEQUENCE (MYELOMA PROTEIN TRO).
RX MEDLINE=76023781; PubMed=809331;
RA Kretzlin H., Altevogt P., Ruban E., Kortt A., Staroscik K.,
RT "The primary structure of a monoclonal IgA-immunoglobulin (IgA Tro.),
RT II. The amino acid sequence of the H-chain, alpha-type, subgroup III;
RT structure of the complete IgA-molecule.";
RL Hoppe-Seyler's Z. Physiol. Chem. 356:1337-1342(1975).
RN [4]
RP DISULFIDE BONDS.
RX MEDLINE=80114124; PubMed=393607;
RA Yang C.-Y., Kretzlin H., Gotz H., Hilschmann N.;
RT "Rule of antibody structure. Primary structure of a human monoclonal
RT IgA-immunoglobulin (myeloma protein Tro). VII. Purification and
RT characterization of the disulfide bridges.";
RL Hoppe-Seyler's Z. Physiol. Chem. 360:1919-1940(1979).
RN [5]
RP REVIEW.
RX MEDLINE=91054387; PubMed=2241915;
RA Kerr M.A.;
RT "The structure and function of human IgA";
RL Biochem. J. 271:285-296(1990).
CC -1- FUNCTION: Ig alpha is the major immunoglobulin class in body
CC secretions. It may serve both to defend against local infection
CC and to prevent access of foreign antigens to the general
CC immunologic system.
CC -1- SUBUNIT: Monomeric or polymeric.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

Dh	18/	SVLPGCAEPNNHCKTFTCTAAVPESTKPLATLTSKS--GNFRPDEVILLPPSEELANLEL	245
Oy	542	VSLTLCVKGYEYPSDIAVWESNQO--PENNYKTTPPLTD-SDG--SPFLYSKLTVDKSRN	596
Dh	246	VTLTCLARGFSPPDVAVVRMVGQSQDELPREKYLTWASRQEPSQGTTFPAVVISILVAEDW	305
Oy	597	QQGVNFCSCVMHEALNHNYQKSLSLSPG-----LDLDETC	632
Dh	306	KKGDTFSCMGHEALPLAFTOKITIDRLAGKPTHVNVSVMVAEVDGTC	352
 RESULT_55			
ID	HVCL_HETFR	STANDARD;	PRT; 370 AA.
AC	P23084;		
DT	01-NOV-1991 (Rel. 20, Created)		
DT	01-NOV-1991 (Rel. 20, Last sequence update)		
DT	10-OCT-2003 (Rel. 42, Last annotation update)		
DE	Ig heavy chain C region (Clone 6125) (Fragment).		
OS	Heterodontus francisci (Horn shark).		
OC	Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Chondrichthyes;		
OC	Elasmobranchii; Galeomorphii; Heterodontidae; Heterodontiformes;		
OX	NCBII_Taxid=7792;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=88328985; Pubmed=3138109;		
RA	Kokubu F., Hinde K., Litman R., Shambloct M.J., Litman G.W.;		
RT	"Complete structure and organization of immunoglobulin heavy chain		
RL	constant region genes in a phylogenetically primitive vertebrate.";		
CC	EMBO J. 7:1979-1988(1988).		
CC	-I- SIMILARITY: Contains 3 immunoglobulin-like domains.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; X07783; CAA30616.1; -.		
DR	PIR; S01851; HVRCSS.		
DR	HSSP; P01857; IRCL.		
DR	InterPro; IPR007110; Ig-like.		
DR	InterPro; IPR003597; Ig_c1.		
DR	InterPro; IPR003006; Ig_MHC.		
DR	Pfam; PF00047; Ig; 3.		
DR	SMART; SMO0407; IGc1; 2.		
DR	PROSITE; PSS0835; IG_LIKE; 3.		
DR	PROSITE; PS00290; IG_MHC; 3.		
FW	Immunoglobulin domain; Immunoglobulin C region; Glycoprotein; Repeat.		
FT	NON TER	1	
FT	DOMAIN	1	
FT	FT	40	IG-LIKE 1.
FT	FT	145	IG-LIKE 2.
FT	DOMAIN	247	IG-LIKE 3.
FT	CARBOHYD	98	(POTENTIAL).
FT	CARBOHYD	132	(POTENTIAL).
FT	CARBOHYD	177	(POTENTIAL).
FT	CARBOHYD	343	(POTENTIAL).
FT	CARBOHYD	347	(POTENTIAL).
FT	CARBOHYD	357	(POTENTIAL).
SO	SEQUENCE	370 AA;	40586 MW; 32833A6EF7ADDED CRC64;
 Query Match 7.9%; Score 270; DB 1; Length 370; Best Local Similarity 30.7%; Pred. No. 3.5e-10; Matches 73; Conservative 43; Mismatches 90; Indels 32; Gaps 9;			
Oy	398	APEPKSCDKHTPCPELLGGPSVFLPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKENRY	457
Dh	136	SPDSSECNM-----PAIKLLPSIEQVL-LEATVTLTVC---VSNAPYGIVNSW-	180
Oy	458	VDGVENVHNAKKPREEQY-----NSTRVSVVLTVLQDMVNGEKYKKNVSNKALLPAIE	512

```

Db 181 -----TOBOKPLSEIAVOPGEDPDSVISTVDISTQAMSEAFVCVSHODLPTLR 233
Oy 513 KTSK-AGGPREPOVYTLPPSRDELTKNQ-VSLTCLVKGYSPDIAMENSGNPPEN-- 568
Db 234 DSIHKAQKDLREBPVSVLPPAEIISAEPLSLTCLVRGSPREIFVKMTVNDKSVNPG 293
Oy 569 NYKTPPVLDSD-GSFFLYSKLTVDSKRWQGNVFCSCVMHEALHNHTOKSLSPG 625
Db 294 NYKTEVVAENDKSSFFIYLSLSTIAEEMASGASISCVVGHAIPLKINTVKNSSG 351

RESULT 56
MUCM ICTPU STANDARD; PRT; 481 AA.
AC P23735;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ig mu chain C region membrane-bound form.
OS Ictalurus punctatus (Channel catfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Siluriformes;
OC Ictaluridae; Ictalurus.
OX NCBI_TaxID=7998;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90384824; PubMed=2119496;
RA Wilson M.R., Marcuz A., van Ginkel F., Miller N.W., Clem L.W.,
RA Middleton D., Warr G.W.;
RT "The immunoglobulin M heavy chain constant region gene of the channel
RT catfish, Ictalurus punctatus: an unusual mRNA splice pattern produces
RT the membrane form of the molecule.";
RL Nucleic Acids Res. 18:5227-5233 (1990).
CC -I- MISCELLANEOUS: During differentiation, B lymphocytes switch from
CC expression of membrane-bound Igm to secretion of Igm. The mu
CC chain of membrane-bound Igm differ in their C-terminal
CC segments.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
CC -----
DR EMBL; X52617; -; NOT_ANNOTATED_CDS.
DR HSBP; P01857; IFC1.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003597; Ig_c1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig; 3.
DR SMART; SM00407; IGC1; 1.
DR PROSITE; PS50835; IG LIKE; 3.
DR PROSITE; PS00290; IG MHC; 2.
KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
KW Transmembrane.
KW NON_TER
FT DOMAIN 1 101 CH1.
FT DOMAIN 102 203 CH2.
FT DOMAIN 204 304 CH3.
FT DOMAIN 305 434 CH4.
FT TRANSMEM 459 478 POTENTIAL.
FT DOMAIN 479 481 CYTOPLASMIC (WITH A LIGHT CHAIN)
FT DISULFID 14 14 INTERCHAIN (BY SIMILARITY).
FT DISULFID 26 85 (BY SIMILARITY).
FT DISULFID 129 188 BY SIMILARITY.
FT DISULFID 233 284 BY SIMILARITY.
FT CARBOHYD 122 122 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

```

FT CARBOHYD 258 258 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 278 278 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 359 359 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 420 420 N-LINKED (GLCNAC. . .) (POTENTIAL).
SO SEQUENCE 481 AA; 53337 MW; 668E784E181C71BA CRC64;

Query Match 7.7%; Score 261.5; DB 1; Length 481;
Best Local Similarity 25.3%; Pred. No. 1.7e-09;
Matches 92; Conservative 60; Mismatches 147; Indels 65; Gaps 16;

Oy 294 SGNLTALAEKGT-----KLHGEVLVVRATQLOKMLTCEVM--GPTSPK 337
Db 78 NGNKKFTCEVNGNGLSKXDAISLQKPEREILHSLTLTPTQTEINDGTATFCLATPSPK 137
Oy 338 LMLSLKLENKAKSKXKEKPEVNVLPNPAAGMQLCSGQVLTLESNIKVLPTWSTPPCP 397
Db 138 -SHTFKWTLEKTDISNRYKEN-IVSQNKGNFTAL-----SVLTLSASEWTSSTSPKCE 189
Oy 398 APEPKSCDKTHT-----CPELLGSPVFLFPPEPKDTLMSRTPEVCVVVDSHEDP 450
Db 190 FQQ-----KHNVPKEASVYAGDTPKQPVKITGSTD-ILIKRAGQLEC---RAEDT 239
Oy 451 EVK-FNNYVDGVEVNAKTKPREEQNSTYRVVSVLT-VLHQMVLNGEKYKCVSNALP 508
Db 240 GFKSIKWLIGNREISLSNL-----SKTVALQTHIGPEEWINGTEFICEVHEAFT 292
Oy 509 APIRK-TISRAKGPPEPOVYTLPPSRDELTKNQVSLTCLVKGYSPDIAMENSGN---- 563
Db 293 QQYKVKVFKENGPEPKVYLLAP--PESSGESVTLTCYKDKDYPPEVAASMLVNDKQV 350
Oy 564 ----GQENNYKTPPVLDSDGSFFLYSKLTVDSKRWQGNVFCSCVMHEALHNHTOKS 619
Db 351 EEVVGYSQN-----TVAIDRNLFVSYSOLITIKTADNNSGVFCLVYHESIKCVKRIIS 406
Oy 620 LSLG 623
Db 407 RSLA 410

RESULT 57
ALC MOUSE STANDARD; PRT; 344 AA.
AC P01878;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ig alpha chain C region.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (MYELOMA ABE48 AND J558).
RX MEDLINE=81261947; PubMed=6790349;
RA Aufferay C., Nageotte R., Sikorav J.-L., Heidmann O., Rougeon F.;
RT "Mouse immunoglobulin A: nucleotide sequence of the structural gene
RT for the alpha heavy chain derived from cloned cDNAs.";
RL Gene 13:365-374 (1981).
RN [2]
RP SEQUENCE OF 1-213 (MOPC 47A).
RX MEDLINE=80049769; PubMed=115869;
RA Robinson E.A., Appella E.;
RT "Amino acid sequence of a mouse myeloma immunoglobulin heavy chain
RT (MOPC 47 A) with a 100-residue deletion.";
RL J. Biol. Chem. 254:11418-11430 (1979).
RN [3]
RP SEQUENCE OF 1-254 AND 291-344 (M511).
RX MEDLINE=81054880; PubMed=6776528;
RA Robinson E.A., Appella E.;
RT "Complete amino acid sequence of a mouse immunoglobulin alpha chain
RT (MOPC 511).";
RL Proc. Natl. Acad. Sci. U.S.A. 77:4909-4913 (1980).
-1- FUNCTION: Ig alpha is the major immunoglobulin class in body

```

```

CC secretions. It may serve both to defend against local infection
CC and to prevent access of foreign antigens to the general
CC immunologic system.
CC -1- MISCELLANEOUS: The final C-region domain is deleted from Ref. 2
CC chain. It was isolated from a myeloma protein that contains 1
CC light and 1 heavy chain per molecule, linked by a disulfide bond.
CC In contrast, normal mouse IgA molecules contain 2 light and 2
CC heavy chains and lack a light-heavy chain disulfide bond.
CC -1- MISCELLANEOUS: M511 chain was isolated from a myeloma protein that
CC binds phosphorylcholine.
CC -1- MISCELLANEOUS: M511 sequence was compared with that of mouse MOPC
CC 47A, and a genetic mechanism for the deletion of the CH3 domain of
CC the mutant chain is proposed.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: D11468; BAA02026.1; ALT_INIT.
CC PIR: A91479; AHMS.
CC HSSP: P01810; 2FBU.
CC InterPro: IPR007110; Ig-like.
CC InterPro: IPR003597; Ig_c1.
CC InterPro: IPR003006; Ig_MHC.
CC Pfam: PF00047; Ig; 3.
CC SMART: SM00407; Igcl; 2.
CC PROSITE: PSS0835; IG_LIKE; 3.
CC PROSITE: PS00290; IG_MHC; 2.
CC Immunoglobulin domain; Immunoglobulin C region; Glycoprotein; Repeat.
CC
CC NON_TER 1 1
CC DOMAIN 6 99 IG-LIKE 1.
CC FT 116 206 IG-LIKE 2.
CC FT 219 321 IG-LIKE 3.
CC FT 26 84 BY SIMILARITY.
CC FT DISULFID 76 100 BY SIMILARITY.
CC FT DISULFID 114 171 BY SIMILARITY.
CC FT DISULFID 138 195 BY SIMILARITY.
CC FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .).
CC FT CARBOHYD 101 101 O-LINKED.
CC FT CARBOHYD 331 331 O-LINKED.
CC FT VARIANT 2 2 S -> A (IN MOPC 47A).
CC FT VARIANT 18 18 S -> C (IN MOPC 47A AND M511).
CC FT VARIANT 67 67 S -> S (IN MOPC 47A AND M511).
CC FT VARIANT 73 73 A -> T (IN MOPC 47A).
CC FT VARIANT 112 112 P -> G (IN M511).
CC FT VARIANT 135 135 S -> Q (IN MOPC 47A AND M511).
CC FT VARIANT 141 141 N -> D (IN MOPC 47A AND M511).
CC FT VARIANT 168 168 Q -> E (IN MOPC 47A).
CC FT VARIANT 212 213 VT -> SQ (IN MOPC 47A).
CC FT VARIANT 235 235 E -> G (IN M511).
CC FT VARIANT 255 255 MISSING (IN M511).
CC FT VARIANT 295 295 T -> D (IN M511).
CC FT VARIANT 301 301 Q -> G (IN M511).
CC FT VARIANT 329 329 N -> Q (IN M511).
CC FT VARIANT 331 331 S -> N (IN M511).
CC SQ SEQUENCE 344 AA; 36876 MW; 3694CFEP9B19A9P8 CRC64;

```

```

QY 391 SNVPCPAPRPKCDKHTTCPELLGSPVFLPFPKXDTMTSRTPVTVVVDVSHEDP 450
DB 101 SPTPTPTPTPTSCQ-----PSLSLQRPALPD-LTSGDASITCTINGLNRPG 148
QY 451 EVKFMVYDGVVHNAKTPREQYNTST---YRVSVTLVLHQDWLNGEKYCKVSNKAL 507
DB 149 AV-FTW-----EPSTGKAVQKKNVQNGCGSVSVLPGCAERNNSGASFCTVTHPE- 201
QY 508 PAPIKTKSKAKQGPREFQVYTLPPSRDELTKNQ-VSLTCLVKGFPYSDIAVEMSGNQ- 565
DB 202 SGLTGLTKATKYVNTVTFPPQVHLLPPESEBLANBELSLTCLVRAVNPKEVLVRWVHGNE 261
QY 566 --PENNYKTPPVYLDSDG--SPELYSKLTVDKSRMOQGVFSCVWHEALHNHYQKSL 621
DB 262 LSPSYLVFEPFLKEPGEATTLVTSVLRVSAETWKQGGQVSCWGHLELPMVFTOKTID 321
QY 622 LSPG 625
DB 322 RLSG 325

```

RESULT 58

```

SHS1 MOUSE STANDARD; PRT; 513 AA.
ID SHS1 MOUSE
AC P97757; O08907; O35924; O08555; O85556; P97796; Q8R559; Q9QX57;
Q9WTN4;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Protein-tyrosine phosphatase non-receptor type subunit 1 (Signal
DE (SHP substrate-1) (SHP-1) (Inhibitory receptor SHPS-1) (Signal-
DE regulatory protein alpha-1) (SIRP-alpha-1) (MSIRP-alpha) (MyD-1
DE antigen) (Brain Ig-like molecule with tyrosine-based activation
DE motifs) (Bit) (p84).
GN PTNS1 OR SHPS1 OR SIRP OR MYD1 OR BIT.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN (1)
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RX TISSUE=Brain;
RX MEDLINE=97223399; PubMed=9070220;
RA Yamao T., Matozaki T., Amano K., Matsuda Y., Takahashi N., Ochi F.,
RA Fujioka Y., Kasuga M.;
RT "Mouse and human SHPS-1: molecular cloning of cDNAs and chromosomal
RT localization of genes.";
RL Biochem. Biophys. Res. Commun. 231:61-67(1997).
RN (2)
RP SEQUENCE FROM N.A. (ISOFORM 2), AND VARIANTS ALA-29; ARG-67; ARG-91;
RP THR-96; SER-128; PRO-194 AND ASN-224.
RX STRAIN=BALB/c; TISSUE=Brain;
RX MEDLINE=97230468; PubMed=9073522;
RA Ohnishi H., Kubota M., Sano S.-I.;
RL "BIT (Bit) maps to mouse chromosome 2.";
RL Genomics 40:504-506(1997).
RN (3)
RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 3), SEQUENCE OF 32-53 AND 422-433
RP (ISOFORM 2), VARIANTS ALA-29; ARG-91; THR-96 AND SER-128,
RP N-GLYCOSYLATION, AND TISSUE SPECIFICITY.
RX STRAIN=BALB/c; TISSUE=Brain, and Cerebellum;
RX MEDLINE=98012243; PubMed=9348339;
RA Comu S., Weng W., Olinick S., Ishwad P., Mi Z., Hempel J., Watkins S.,
RA Lagenaar C.F., Narayanan V.;
RT "The murine p84 neural adhesion molecule is SHPS-1, a member of the
RT phosphatase-binding protein family.";
RL J. Neurosci. 17:8702-8710(1997).
RN (4)
RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 3), VARIANTS ALA-29; ARG-67;
RP PRO-74; ALA-83; 86-TYR-VAL-87; ILE-90; ARG-91; THR-96; ALA-114;
RP ILE-118; SER-128; PRO-194; ASN-224; PHE-351 AND ASP-365,
RP N-GLYCOSYLATION, PHOSPHORYLATION ON TYROSINE RESIDUES, INTERACTION

```

RP WITH PTPN6, AND TISSUE SPECIFICITY.
RC STRAIN=C57BL/6; TISSUE=fetal thymus;
RX MEDLINE=96380500; PubMed=9712903;
RA Veilleux A., Thibaudan E., Latour S.;
RT "High expression of inhibitory receptor SHP-1 and its association
RT with protein tyrosine phosphatase SHP-1 in macrophages.";
RL J. Biol. Chem. 273:22719-22728(1998).
[5]
RP SEQUENCE FROM N.A. (ISOFORMS 1, 2 AND 3), VARIANTS ALA-29; ARG-67;
RP PRO-74; ALA-63; 86-TYR-VAL-87; ILE-90; ARG-91; THR-96; ALA-114;
RP ILE-118; SER-128; PRO-194; ASN-224; PHE-351 AND ASP-365, AND
RP N-GLYCOSYLATION.
RC STRAIN=129/SvJ, and C57BL/6; TISSUE=Brain, and Liver;
RX MEDLINE=20053880; PubMed=10585853;
RA Sano S.-I., Ohnishi H., Kubota M.;
RT "Gene structure of mouse BTK/SHPS-1.";
RL Biochem. J. 344:667-675(1999).
[6]
RP SEQUENCE FROM N.A. (ISOFORM 2), AND VARIANTS ALA-29; ARG-67;
RP PRO-74; ALA-63; 86-TYR-VAL-87; ILE-90; ARG-91; THR-96; ALA-114;
RP ILE-118; SER-128; PRO-194; ASN-224; PHE-351 AND ASP-365.
RC STRAIN=129/SvJ, and ISS;
RX MEDLINE=21363810; PubMed=11471062;
RA Ehlinger M.A., Thompson J., Conroy O., Xu Y., Yang F., Canfield J.,
RA Beeson M., Gordon L., Bennett B., Johnson T.E., Sikele J.M.;
RT "High-throughput sequence identification of gene coding variants
RT within alcohol-related OTUs.";
RL Mamm. Genome 12:657-663(2001).
[7]
RN SEQUENCE FROM N.A. (ISOFORM 1), AND VARIANT PRO-194.
RP TISSUE=Placenta;
RC TISSUE=Placenta;
RA Wang H., Chen Z., Ullrich A.;
RT "Epidermal growth factor-induced association of SHP2 with mouse SIRP-
RT alpha.";
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
[8]
RP FUNCTION, TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.
RX MEDLINE=90152134; PubMed=2203162;
RA Chuang W., Laganaur C.F.;
RT "Central nervous system antigen P84 can serve as a substrate for
RT neurite outgrowth.";
RL Dev. Biol. 137:219-232(1990).
[9]
RN N-GLYCOSYLATION, PHOSPHORYLATION BY JAK2 IN RESPONSE TO GROWTH
RP HORMONE, AND INTERACTIONS WITH JAK2 AND PTPN11.
RX MEDLINE=96175985; PubMed=9507023;
RA Stofega M.R., Wang H., Ullrich A., Carter-Su C.;
RT "Growth hormone regulation of SIRP and SHP-2 tyrosyl phosphorylation
RT and association.";
RL J. Biol. Chem. 273:7112-7117(1998).
[10]
RP INTERACTION WITH CD47, AND TISSUE SPECIFICITY.
RX MEDLINE=99091586; PubMed=9872987;
RA Jiang P., Laganaur C.F., Narayanan V.;
RT "Integrin-associated protein is a ligand for the P84 neural adhesion
RT molecule.";
RL J. Biol. Chem. 274:559-562(1999).
[11]
RN INTERACTION WITH FGR.
RX MEDLINE=20130295; PubMed=10662797;
RA Gresham H.D., Dale B.M., Potter J.W., Chang P.W., Vines C.M.,
RA Lowell C.A., Laganaur C.F., Willman C.L.;
RT "Negative regulation of phagocytosis in murine macrophages by the Src
RT kinase family member, Fgr.";
RL J. Exp. Med. 191:515-528(2000).
-1- FUNCTION: Immunoglobulin-like cell surface receptor for CD47. Acts
as docking protein and induces translocation of PTPN6,
PPN1 and other binding partners from the cytosol to the
plasma membrane. Supports adhesion of cerebellar neurons, neurite
outgrowth and glial cell attachment. May play a key role in
intracellular signaling during synaptogenesis and in synaptic
function. Involved in the negative regulation of receptor tyrosine
kinase-coupled cellular responses induced by cell adhesion, growth

CC factors or insulin. Mediates negative regulation of phagocytosis,
CC mast cell activation and dendritic cell activation. CD47 binding
CC prevents maturation of immature dendritic cells and inhibits
CC cytokine production by mature dendritic cells (By similarity).
CC -1- SUBUNIT: Binds PTPN11 when tyrosine-phosphorylated, except in
CC macrophages, where it primarily binds PTPN6. Binds GRB2 vitro.
CC Binds FGR. Binds JAK2 irrespective of its phosphorylation status
CC and forms a stable complex. Binds SCAP1 and/or SCAP2. The
CC resulting complex recruits FVB. Binds PTK2B (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC Event=Alternative splicing; Named isoforms=3;
CC Comment=Additional isoforms seem to exist;
CC Name=1; Synonyms=a;
CC IsoId=P97797-1; Sequence=Displayed;
CC Name=2; Synonyms=a', Large;
CC IsoId=P97797-2; Sequence=VSP_007032;
CC Name=3; Synonyms=b, Small;
CC IsoId=P97797-3; Sequence=VSP_007031;
CC -1- TISSUE SPECIFICITY: Highly expressed in cerebral cortex, brain,
CC spinal cord, cerebellum and spleen, and at much lower levels in
CC kidney, thymus, heart, lung and liver. Within the cerebellum,
CC highly expressed throughout the molecular layer, and in synaptic
CC glomeruli in the granule cell layer. Detected in neurons of the
CC hippocampus and dentate gyrus, and in olfactory bulb. Not detected
CC in Purkinje cells. Highly expressed in the plexiform layers, optic
CC fiber layer and the outer segments of the photoreceptor layer in
CC the retina. Highly expressed in macrophages. Isoform 3 is detected
CC at very low levels in all tissues tested.
CC -1- DEVELOPMENTAL STAGE: Highly expressed in the CNS of embryos from
CC day 7 to 17.
CC -1- PTM: N-glycosylated.
CC -1- PTM: Phosphorylated on tyrosine residues.
CC -1- SIMILARITY: Contains 2 immunoglobulin-like C1-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL, D87967; BA13520.1; -;
DR EMBL, D87968; BA13521.1; -;
DR EMBL, D85785; BA20376.1; -;
DR EMBL, U89694; AAB92591.1; -;
DR EMBL, AF072543; AAC24886.1; -;
DR EMBL, AF072544; AAC24887.1; -;
DR EMBL, AB024507; BAA89290.1; -;
DR EMBL, AB024507; BAA89290.1; JOINED.
DR EMBL, AB024501; BAA89289.1; JOINED.
DR EMBL, AB024501; BAA89289.1; JOINED.
DR EMBL, AB024502; BAA89290.1; JOINED.
DR EMBL, AB024502; BAA89290.1; JOINED.
DR EMBL, AB024503; BAA89290.1; JOINED.
DR EMBL, AB024504; BAA89290.1; JOINED.
DR EMBL, AB024505; BAA89290.1; JOINED.
DR EMBL, AB024506; BAA89289.1; JOINED.
DR EMBL, AB024507; BAA89289.1; -;
DR EMBL, AB024507; BAA89289.1; JOINED.
DR EMBL, AB024506; BAA89289.1; JOINED.
DR EMBL, AB024501; BAA89289.1; JOINED.
DR EMBL, AB024502; BAA89289.1; JOINED.
DR EMBL, AB024503; BAA89289.1; JOINED.
DR EMBL, AB024504; BAA89289.1; JOINED.
DR EMBL, AB024505; BAA89289.1; JOINED.
DR EMBL, AB024506; BAA89289.1; JOINED.
DR EMBL, AF332079; AAK56108.1; -;
DR EMBL, AF332080; AAK56108.1; -;
DR EMBL, Y10349; CA71375.1; -;
DR PIR, JCS289; JCS289.
DR MGI, MGI:108563; Pcpn1.
CO, GO:0008580; F:cytoskeletal regulator activity; IMP.

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; X62515; CAA44373.1; -

DR EMBL; M65289; AAB52700.1; -

DR EMBL; AL445795; CAC18534.1; -

DR EMBL; M64283; AAB52699.1; -

DR EMBL; S76436; AAB21121.2; -

DR EMBL; L22078; -; NOT_ANNOTATED_CDS.

DR PIR; A38096; A38096.

DR HSP; P00740; IEDM.

DR Sigena-2DPAGE; P96160; -

DR Genew; HGNC:5273; HSPG2.

DR MIM; 142461; -

DR MIM; 255800; -

DR InterPro; IPR008985; Cona_like_1ec_g1.

DR InterPro; IPR000742; EGF_2.

DR InterPro; IPR006209; EGF_like.

DR InterPro; IPR006210; IEGF.

DR InterPro; IPR007110; IG-like.

DR InterPro; IPR003599; IG_c2.

DR InterPro; IPR003598; IG_c2.

DR InterPro; IPR003596; IG_v.

DR InterPro; IPR000034; Laminin_B.

DR InterPro; IPR002049; Laminin_EGF.

DR InterPro; IPR001791; Laminin_G.

DR InterPro; IPR002172; LDL_receptor_A.

DR InterPro; IPR000082; SEA_domain.

DR Pfam; PF00008; EGF; 4.

DR Pfam; PF00047; IG; 22.

DR Pfam; PF00052; laminin_B; 3.

DR Pfam; PF00053; laminin_EGF; 7.

DR Pfam; PF00054; laminin_G; 3.

DR Pfam; PF00057; ldl_recept_a; 4.

DR Pfam; PF01390; SEA; 1.

DR PRINTS; PR00261; LDLRECEPTOR.

DR PRODOM; PD003031; Laminin_B; 3.

DR SMART; SM00181; EGF; 15.

DR SMART; SM00180; EGF_Lam; 12.

DR SMART; SM00409; IG; 22.

DR SMART; SM00408; IGC2; 21.

DR SMART; SM00406; IGV; 7.

DR SMART; SM00281; LamB; 3.

DR SMART; SM00282; LamG; 3.

DR SMART; SM00182; LDLa; 4.

DR SMART; SM00200; SEA; 1.

DR PROSITE; PS00022; EGF_1; 9.

DR PROSITE; PS01186; EGF_2; 6.

DR PROSITE; PS00026; EGF_3; 4.

DR PROSITE; PS00835; IG_Like; 22.

DR PROSITE; PS50025; LAM_G_DOMAIN; 3.

DR PROSITE; PS01248; LAMININ_TYPE_EGF; 11.

DR PROSITE; PS01209; LDLRA_1; 4.

DR PROSITE; PS00068; LDLRA_2; 4.

DR PROSITE; PS50024; SEA; 1.

KM Signal; Basement membrane; Proteoglycan; Repeat; Glycoprotein; Heparan sulfate; Laminin EGF-like domain; Immunoglobulin domain; Extracellular matrix; EGF-like domain; Disease mutation.

KW

FT SIGNAL 1 21

FT CHAIN 22 4391

FT DOMAIN 80 194

FT DOMAIN 198 235

FT DOMAIN 284 320

FT DOMAIN 324 360

FT DOMAIN 367 404

FT DOMAIN 405 504

FT DOMAIN 521 530

FT DOMAIN 531 730

FT DOMAIN 731 763

FT DOMAIN 764 813

FT DOMAIN 814 871

FT DOMAIN 879 923

BASEMENT MEMBRANE-SPECIFIC HEPARAN SULFATE PROTEOGLYCAN CORE PROTEIN.

SEA.

LDL-RECEPTOR CLASS A 1.

LDL-RECEPTOR CLASS A 2.

LDL-RECEPTOR CLASS A 3.

LDL-RECEPTOR CLASS A 4.

IG-LIKE C2-TYPE 1.

LAMININ EGF-LIKE 1 (N-TERMINAL).

LAMININ DOMAIN IV 1 (DOMAIN III A).

LAMININ EGF-LIKE 2.

LAMININ EGF-LIKE 3.

LAMININ EGF-LIKE 4 (INCOMPLETE).

FT DOMAIN 924 933 LAMININ EGF-LIKE 5 (N-TERMINAL).

FT DOMAIN 934 1125 LAMININ DOMAIN IV 2 (DOMAIN III B).

FT DOMAIN 1126 1158 LAMININ EGF-LIKE 5 (C-TERMINAL).

FT DOMAIN 1159 1208 LAMININ EGF-LIKE 6.

FT DOMAIN 1209 1265 LAMININ EGF-LIKE 7.

FT DOMAIN 1275 1324 LAMININ EGF-LIKE 8.

FT DOMAIN 1325 1334 LAMININ EGF-LIKE 9 (N-TERMINAL).

FT DOMAIN 1335 1529 LAMININ EGF-LIKE 9 (N-TERMINAL).

FT DOMAIN 1530 1562 LAMININ EGF-LIKE 9 (C-TERMINAL).

FT DOMAIN 1563 1612 LAMININ EGF-LIKE 10.

FT DOMAIN 1613 1670 LAMININ EGF-LIKE 11.

FT DOMAIN 1671 1771 LAMININ EGF-LIKE 12.

FT DOMAIN 1772 1865 LAMININ EGF-LIKE 13.

FT DOMAIN 1866 1955 LAMININ EGF-LIKE 14.

FT DOMAIN 1956 2051 LAMININ EGF-LIKE 15.

FT DOMAIN 2052 2151 LAMININ EGF-LIKE 16.

FT DOMAIN 2152 2244 LAMININ EGF-LIKE 17.

FT DOMAIN 2245 2340 LAMININ EGF-LIKE 18.

FT DOMAIN 2341 2436 LAMININ EGF-LIKE 19.

FT DOMAIN 2437 2533 LAMININ EGF-LIKE 20.

FT DOMAIN 2534 2629 LAMININ EGF-LIKE 21.

FT DOMAIN 2630 2726 LAMININ EGF-LIKE 22.

Query Match 5.4%; Score 184; DB 1; Length 4391;

Best Local Similarity 20.4%; Pred. No. 0.0026;

Matches 156; Conservative 97; Mismatches 255; Indels 216; Gaps 37;

5 VPRHLLVQLALPRA-ATQGNKV--VLAKGD-----TVELCTASQKKSIOFHW 53

2477 LPARHQVHGRRLRLQTPADSGEYVCRVWGSGTQASVLTITQGLSGSHSGVAV-- 2534

54 KNSQIILNQGSLFKGPKSKNDKADSKRS--LMDQ--GNP-----LITKL 99

2535 ----PVRISSSASLANGHFTLNCVLSQAHPHTTWYKKGSSLPSSHQIVGSRLLRPQV 2590

100 KIEDSDTYICEVED--OKEEVOLLVFLTANSPTHLAQO--SLTLLESP--GSSPS 152

2591 TPADSGEYVCRVWGSGTQASVLTITQGLSGSHSGVAV--LITKL 2640

153 V-----OC--RSP-----RKNIGQ-----GKTLVSQLELADSGTCTVLO 188

2641 VVEGQTLNLCVAVRQALITITWYKKGSSLPSSHQIVGSRLLRPQV 2700

189 NQKKVEKIDIV-----LAPKASSIVYKKGGEVEFPPLAFYTEKLTGS 235

2701 NIDALEASIVISVPSAGSPASGSPAPRIRIESSSHVAGETLIDL-----CVVQQA 2755

236 GELMWQAEKASSSKSWITPFLNKKKEVSKRVTDQPKLQWQKPLPHLTLPOALPOVAGSG 295

2756 AQTVMHKKGGSLPSHHOT-----RGRRLNHHVSP-----ADSG 2789

296 NLTLALEAKTKLHOEVLNVMRATOLQ-----NLTC 328

2790 EYVCRVWGSGPLLEASV-LVTIESSGSAHVAPGAPRIRIESSSRVABQOTLIDLC 2848

329 EYWGSPSPKMLSLKLENKAKVSKKEKPVVTLN-----PAGMQLSLSGOVLLEST 384

2849 VVEGQALVQVTWKRGGNLPLAR-HOVHGPLRLRNQVSPADSGEYSCVYTSSTG-LEASV 2906

385 KVLPTWSTPVPCAPPEKSKDKNHTCELLGSGSVLFPKPKDITLMSITPEVTCVVD 444

2907 LVITPESSPPIPP-----GLAQPIITASSSHVTEGQTLIDLCVVG 2950

445 VSHEDPEVKMNVYDVEVNAKTKPREBOYNSTYRVSVLTVLHODMLNGEYKCVSN 504

2951 QAH-----AQTVMHKKG-----GSLPARHQVHGRRLNHHVSP-----DSGEYVCRAS 2995

505 KALP---APIEKITSKAKQP---REPOVTLTPSSDELTKNOVSLTCLV-KKPYPSDIA 557

2996 GPGEBOEASTVTVVPSEGSYRLRSFVLSIDPSSTVQGGDASFKCLIHDAAP--IS 3053

558 VEMESNQF--ENNVKTPPLVLDSDGFFLXSKLTVDKSRNQGNVSC-----SV 606

```

DB 3054 LEWKTNOELEDVHISP-----NGSI-----ITIVGTRPSNNGTCVAVSNAYGAQSV 3103
QY 607 MHEBALNHNTQKSLSLSP-----GLQIDETCAEAGQELDLMT 645
DB 3104 VNLSVHGPPPTVSVPKGVKAVTLECVASAGEPRRSARMT 3147

RESULT 60
LACT_MOUSE
ID LACT_MOUSE STANDARD; PRT; 105 AA.
AC P01843;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ig lambda-1 chain C region.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83014953; PubMed=6812053;
RA Seising E., Miller J., Wilson R., Storb U.;
RT "Evolution of mouse immunoglobulin lambda genes.";
RL Proc. Natl. Acad. Sci. U.S.A. 79:4681-4685(1982).
RN [2]
RP SEQUENCE FROM N.A. (MOPC 315).
RX MEDLINE=8148806; PubMed=6259534;
RA Bothwell A.L.M., Paskind M., Schwartz R.C., Sonenshein G.E.,
RA Gelfer M.L., Baltimore D.;
RT "Dual expression of lambda genes in the MOPC-315 plasmacytoma.";
RL Nature 290:65-67(1981).
RN [3]
RP SEQUENCE FROM N.A. (S43).
RX MEDLINE=8220143; PubMed=6283385;
RA Bothwell A.L.M., Paskind M., Rech M., Imanishi-Kari T., Rajewsky K.,
RA Baltimore D.;
RT "Somatic variants of murine immunoglobulin lambda light chains.";
RL Nature 298:380-382(1982).
RN [4]
RP SEQUENCE (MYELOMA PROTEIN MOPC 104E).
RX MEDLINE=71107854; PubMed=5276767;
RA Appella E.;
RT "Amino acid sequences of two mouse immunoglobulin lambda chains.";
RL Proc. Natl. Acad. Sci. U.S.A. 68:590-594(1971).
CC -1- MISCELLANEOUS: The MOPC 315 cell line produces 2 light chains, 1
CC normal lambda-2 chain and 1 abnormal lambda-1 chain that is
CC missing a large part of the V region. The C region sequence (shown
CC here) appears completely normal.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC -----
CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; J00582; AAA51636.1; -
CC DR EMBL; J00587; AAS59672.1; -
CC DR PIR; A93922; LIMS.
CC DR PDB; 1JNH; 06-FEB-02.
CC DR InterPro; IPR007110; Ig-1like.
CC DR InterPro; IPR003597; Ig cl.
CC DR InterPro; IPR003006; Ig_MHC.
CC DR Pfam; PF00047; Ig_1.
CC DR SMART; SM00407; IGL1; 1.
CC DR PROSITE; PSS0835; IG_LIKE; 1.
CC DR PROSITE; PSS00290; IG_MHC; 1.
CC Immunoglobulin domain; Immunoglobulin C region; 3D-structure.
CC FT NON_TER 1 1
CC DOMAIN 6 100 IG-LIKE.

```

```

FT DISULFID 27 86
FT DISULFID 104 104 INTERCHAIN (WITH HEAVY CHAIN).
FT CONFLICT 19 20 ET -> TE (IN REF. 4).
FT CONFLICT 56 56 Q -> E (IN REF. 4).
FT CONFLICT 75 75 MISSING (IN REF. 4).
FT CONFLICT 81 82 HS -> SH (IN REF. 4).
FT CONFLICT 85 85 S -> SS (IN REF. 4).
FT CONFLICT 96 96 E -> O (IN REF. 4).
SO SEQUENCE 105 AA; 11575 MW; A89P2B09BECFA018 CRC64;

Query Match 5.2%; Score 177.5; DB 1; Length 105;
Best Local Similarity 35.9%; Pred. No. 5.2e-05;
Matches 37; Conservative 18; Mismatches 43; Indels 5; Gaps 3;

QY 521 QPR-EPQVYTLPPSRDELTKNOVSLTCLVKGPYPSDIAVWESNQP-ENNYKTPPVLD 578
DB 1 QKSSPSVTLFPSSSELETNATVLTCTIDPYGVVYDMKVDGPTVQGMETTPSKQ 60
QY 579 SDGSFPLYSKLTVDKSRMOCNVFSCSVHMEALHNHYTQKSIS 621
DB 61 SNKKYVASSYLTUTARAMERHSSYCOVTHE---GHYVKSIS 100

RESULT 61
PGBM_MOUSE
ID PGBM_MOUSE STANDARD; PRT; 3707 AA.
AC Q05793;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Basement membrane-specific heparan sulfate proteoglycan core
DE protein precursor (HSPG) (Perlecan) (PLC).
GN HSPG2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Melanoma;
RX MEDLINE=92078153; PubMed=1744087;
RA Noonan D.M., Fulle A., Valente P., Cai S., Horigan E., Sasaki M.,
RA Yamada Y., Hassell J.R.;
RT "The complete sequence of perlecan, a basement membrane heparan
RT sulfate proteoglycan, reveals extensive similarity with laminin A
RT chain, low density lipoprotein-receptor, and the neural cell adhesion
RT molecule.";
RL J. Biol. Chem. 266:22939-22947(1991).
RN [2]
RP SEQUENCE OF 940-1601 AND 1870-2600 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=8904110; PubMed=2972708;
RA Noonan D.M., Horigan E.A., Ledbetter S.R., Vogel G., Sasaki M.,
RA Yamada Y., Hassell J.R.;
RT "Identification of cDNA clones encoding different domains of the
RT basement membrane heparan sulfate proteoglycan.";
RL J. Biol. Chem. 263:16379-16387(1988).
CC -1- FUNCTION: This protein is an integral component of basement
CC membranes. It is responsible for the fixed negative electrostatic
CC charge and is involved in the charge-selective ultrafiltration
CC properties. It serves as an attachment substrate for cells.
CC -1- SUBUNIT: Purified perlecan has a strong tendency to aggregate in
CC dimers or stellate structures. It interacts with other basement
CC membrane components such as laminin, prolargin and collagen type
CC IV.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: Found in the basement membrane.
CC -1- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
CC AND O-LINKED OLIGOSACCHARIDES.
CC -1- SIMILARITY: Contains 4 LDL-receptor class A domains.
CC -1- SIMILARITY: Contains 11 laminin EGF-like domains.
CC -1- SIMILARITY: Contains 3 laminin IV domains.
CC -1- SIMILARITY: Contains 15 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 3 laminin G-like domains.

```



```

FT DISULFID 1792 1839 BY SIMILARITY.
FT DISULFID 1886 1932 BY SIMILARITY.
FT DISULFID 1976 2021 BY SIMILARITY.
FT DISULFID 2073 2118 BY SIMILARITY.
FT DISULFID 2170 2215 BY SIMILARITY.
FT DISULFID 2268 2313 BY SIMILARITY.

Query Match 5.2%; Score 177.5; DB 1; Length 3707;
Best Local Similarity 17.7%; Pred. No. 0.0055; Index 193; Gaps 25;
Matches 116; Conservative 94; Mismatches 251;

QY 33 KGDVETLTCTASQKSIQFHMKNNSQIKLGNQSFLLTKGPSKLNDRADSRSLMDQGNF 92
1968 EGRTVLVCRAAGVPSASITWRKEG-----GSLFRFQAHSSR----- 2005
QY 93 PLTIKUKLKEDSDTYIC---EVEDQKEVQLIVFGLT-----ANSTHLL 134
2006 -LRLHMSVADSGEYVCRRANNNDIAQETSIMISVSSTNSPPAPASAPARISSSSRYVA 2064
QY 135 QGGSLTLTLESPPGSSPSVQCRSPRGKNI-----QGGKTLVSQLELDQSGTWTCTVLQN 189
2065 EGGTLTLCVVPGHAAHQVTMHR-KGGSLPTHQTHGSRRLRYQVSSADSGEYVCSVLSS 2123
QY 190 OKKVEFKIDIVLAFQKASSIVYKKEGEVFEFPLAFVTEKLTGSGELMWQAEARASSSK 249
2124 SGPLEASVLVSI-----PAAANV----- 2143
QY 250 SWITFDLKNKEVSKRVATODPKIQMGKLEPLHLTLPOALFOYAGSGLTLAEAKTGK 309
2144 -----IPGVVPIRIETSSSRVAEGOTLDLSCVVP-----GQAH 2177
QY 310 QEVNIVMVRATQIQKMLTCEVMGPTSPKMLSLKLEKKAKEKREKVPVVLNPEAGMG 369
2178 AQV-----TMHRGSSLPAGHQVGHMLRLNR-----VSPADSGEYS 2214
QY 370 CLSDSGQVLLNESNIKVLPTWSTPVPPCAPREPKSCDKTHCPRLGSPVFLPPPKD 429
2215 CQVYTGSSGT-LEASVLTITASESPRIAP-----GLAQPVYIESSSH 2257
QY 430 LMTSRPEVTCVVDVSHDEPEVKFMYVDGVEHNAKTPREQYNSTYRVSVLTVLH 489
2258 LTGEGQVLDLKCVPFGQH-----AQVTHMKRG-----SLPARHQTHGSLRLYQLSPA-- 2306
QY 490 QDMLENGEKYCKXSNKLP---APIETISKAGQP---REPQVTLTPRSDELITKNQVS 543
2307 ----DSGEYVCQVAGSSHPHEASFKLTVPSQNSFRILSPVISIEPSTVQOQGDAS 2362
QY 544 LTCLV-KGFYPSDIAVWESNGQP-ENNYKTPPVLDSDGSPFLY--SKLTVDKSRWQ 598
2363 FKLILHEGAMP--IKYEMKIRDOLEEDNVHISF-----NSIITIVAPGPAIWEPTACVA 2415
QY 599 GNVFSC--SYMHEALNHNYTQKSLISF-----GLQIDETCAADQELDGLMT 645
2416 SNVYGMAGQSVNLVSHGPTVSVLPBEGPVHVKKQDITTECISGEBRSSPRT 2469

RESULT 62
VCAI_VCAI RAT STANDARD; PRT; 739 AA.
AC P29534;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Vascular cell adhesion protein 1 precursor (V-CAM 1).
GN VCAI OR VCAI-1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
MEDLINE=92181437; PubMed=1371918;

```

```

RA Hession C., Moy P., Tizard R., Chisholm P., Williams C., Wysk M.,
RA Burkle L., Miyake K., Kincaid P., Lobb R.;
RT "Cloning of murine and rat vascular cell adhesion molecule-1.";
RL Biochem. Biophys. Res. Commun. 183:163-169(1992).
CC -1- FUNCTION: IMPORTANT IN CELL-CELL RECOGNITION. APPEARS TO FUNCTION
CC IN LEUKOCYTE-ENDOTHELIAL CELL ADHESION. INTERACTS WITH THE BETA-1
CC INTEGRIN VIA4 ON LEUKOCYTES, AND MEDIATES BOTH ADHESION AND SIGNAL
CC TRANSDUCTION. THE VCAI/VIA4 INTERACTION MAY PLAY A
CC PATHOPHYSIOLOGIC ROLE BOTH IN IMMUNE RESPONSES AND IN LEUKOCYTE
CC EMIGRATION TO SITES OF INFLAMMATION.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: Expressed on inflamed vascular endothelium, as
CC well as on macrophage-like and dendritic cell types in both normal
CC and inflamed tissue.
CC -1- SIMILARITY: Contains 7 immunoglobulin-like C2-type domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; M84488; AAA42332.1; -.
CC PIR; J50675; J50675.
CC HSSP; P19320; 1VCA.
CC InterPro; IPR003987; ICM VCAI-1.
CC InterPro; IPR007110; I9-11ke.
CC InterPro; IPR003598; I9 C2.
CC InterPro; IPR003989; VCAI-1.
CC Pfam; PF00047; I9; 5.
CC DR PRINTS; PRO1472; ICAVCAI.
CC DR PRINTS; PRO1474; VCAI.
CC SMART; SM00408; I9C2; 3.
CC PROSITE; PS50835; I9 LIKE; 5.
CC Immunoglobulin domain; Glycoprotein; Cell adhesion; Transmembrane;
CC Repeat; Signal.
CC
CC FT SIGNAL 1 24 PROBABLE.
CC FT CHAIN 25 739 VASCULAR CELL ADHESION PROTEIN 1.
CC FT DOMAIN 25 698 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 699 720 POTENTIAL.
CC FT DOMAIN 721 739 CYTOPLASMIC (POTENTIAL).
CC FT DOMAIN 25 111 IG-LIKE C2-TYPE 1.
CC FT DOMAIN 119 212 IG-LIKE C2-TYPE 2.
CC FT DOMAIN 223 309 IG-LIKE C2-TYPE 3.
CC FT DOMAIN 312 397 IG-LIKE C2-TYPE 4.
CC FT DOMAIN 408 506 IG-LIKE C2-TYPE 5.
CC FT DOMAIN 514 595 IG-LIKE C2-TYPE 6.
CC FT DOMAIN 601 682 IG-LIKE C2-TYPE 7.
CC FT DISULFID 47 95 BY SIMILARITY.
CC FT DISULFID 52 99 BY SIMILARITY.
CC FT DISULFID 137 195 BY SIMILARITY.
CC FT CARBOHYD 273 273 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 424 424 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 531 531 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 561 561 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 650 650 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 650 650 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SQ SEQUENCE 739 AA; 81246 MW; 560885A1A1B100C CRC64;

Query Match 5.1%; Score 174.5; DB 1; Length 739;
Best Local Similarity 20.4%; Pred. No. 0.001; Index 221; Gaps 29;
Matches 125; Conservative 81; Mismatches 185;

QY 33 KGDVETLTCTASQKSIQFHMKNNSQIKLGNQSFLLTKGPSKLNDRADSRSLMDQGNF 92
238 EGAAYVMTCASGLPAPFLFMSK-----LDNGVQLL-----SGNA 274
QY 93 PLTIKUKLKEDSDTYIC---EVEDQKEVQLIVFGLT-----ANSTHLL 134
275 TLTLLAMRMBDSGIYCEGVNLVGRDKTEVELVQKRPFTVDISPSQVAAQVGDSSVLT 334
QY 143 LBSPPGSSPSVQCRSPRGKNIQ-----GKTLVSQLELDQSGTWTCTVLQNKKVEF 195

```



```

FT DOMAIN 3482 3572 IG-LIKE C2-TYPE 22.
FT DOMAIN 3580 3667 IG-LIKE C2-TYPE 23.
FT DOMAIN 3686 3777 IG-LIKE C2-TYPE 24.
FT DOMAIN 3817 3908 IG-LIKE C2-TYPE 25.
FT DOMAIN 3920 4009 IG-LIKE C2-TYPE 26.
FT DOMAIN 4018 4106 IG-LIKE C2-TYPE 27.
FT DOMAIN 4109 4201 IG-LIKE C2-TYPE 28.
FT DOMAIN 4212 4297 IG-LIKE C2-TYPE 29.
FT DOMAIN 4302 4387 IG-LIKE C2-TYPE 30.
FT DOMAIN 4400 4485 IG-LIKE C2-TYPE 31.
FT DOMAIN 4489 4580 IG-LIKE C2-TYPE 32.
FT DOMAIN 4588 4678 IG-LIKE C2-TYPE 33.
FT DOMAIN 4681 4771 IG-LIKE C2-TYPE 34.
FT DOMAIN 4873 4961 IG-LIKE C2-TYPE 35.
FT DOMAIN 4965 5057 IG-LIKE C2-TYPE 36.
FT DOMAIN 5067 5160 IG-LIKE C2-TYPE 37.
FT DOMAIN 5171 5260 IG-LIKE C2-TYPE 38.
FT DOMAIN 5277 5366 IG-LIKE C2-TYPE 39.
FT DOMAIN 5383 5472 IG-LIKE C2-TYPE 40.
FT DOMAIN 5487 5578 IG-LIKE C2-TYPE 41.
FT DOMAIN 5595 5685 IG-LIKE C2-TYPE 42.
FT DOMAIN 5701 5790 IG-LIKE C2-TYPE 43.
FT DOMAIN 5815 5904 IG-LIKE C2-TYPE 44.
FT DOMAIN 5925 6014 IG-LIKE C2-TYPE 45.
FT DOMAIN 6038 6130 IG-LIKE C2-TYPE 46.
FT DOMAIN 6150 6239 IG-LIKE C2-TYPE 47.
FT DOMAIN 6275 6368 FIBROECTIN TYPE-III.
FT DOMAIN 6413 6502 IG-LIKE C2-TYPE 48.
FT DOMAIN 6507 6596 IG-LIKE C2-TYPE 49.
FT DISULFID 568 621 POTENTIAL.
FT DISULFID 2908 2975 POTENTIAL.
FT DISULFID 3015 3065 POTENTIAL.
FT DISULFID 3707 3759 POTENTIAL.
FT DISULFID 3826 3890 POTENTIAL.
FT DISULFID 5092 5157 POTENTIAL.
FT DISULFID 5298 5350 POTENTIAL.
FT DISULFID 5508 5560 POTENTIAL.
FT DISULFID 5616 5669 POTENTIAL.
FT DISULFID 5722 5764 POTENTIAL.
FT DISULFID 5836 5901 POTENTIAL.
FT DISULFID 5946 5998 POTENTIAL.
FT DISULFID 6036 6171 POTENTIAL.
FT DISULFID 6421 6486 POTENTIAL.
FT CONFLICT 2137 2137 A -> P (IN REF. 1).
FT CONFLICT 2245 2247 AKA -> PKP (IN REF. 1).
FT CONFLICT 2258 2258 A -> P (IN REF. 1).
FT CONFLICT 2284 2284 E -> G (IN REF. 1).
FT CONFLICT 2297 2297 M -> I (IN REF. 1).
FT CONFLICT 3531 3531 A -> G (IN REF. 1).
FT CONFLICT 3884 3888 DAGEV -> RRRRI (IN REF. 1).
FT CONFLICT 3929 3929 A -> V (IN REF. 1).
FT CONFLICT 5134 5134 A -> P (IN REF. 1).
FT CONFLICT 5145 5145 T -> S (IN REF. 1).
FT CONFLICT 5185 5185 G -> A (IN REF. 1).
FT CONFLICT 5199 5199 K -> N (IN REF. 1).
FT CONFLICT 5202 5202 L -> F (IN REF. 1).
FT CONFLICT 5213 5213 F -> L (IN REF. 1).
FT CONFLICT 6178 6178 A -> G (IN REF. 1).
FT CONFLICT 6268 6268 K -> E (IN REF. 1).
SQ SEQUENCE 6632 AA; 731665 MW; 262d3bDD62960E89 CRC64;

Query Match 5.1%; Score 174; DB 1; Length 6632;
Best Local Similarity 20.7%; Pred. No. 0.02; Indels 170; Gaps 24;
Matches 132; Conservative 82; Mismatches 255;

QY 20 PAATQGNKVVIGKGDVETLTASQKSIQFHWKNSNQIKLIGNOGSFLTKGPKSLNDR 79
Db 3580 PLTVQGIKRYEVEVGKRAELVAVVEGKPEVVKMFQDGVIAIDNGVIRKKG-----3632
QY 80 ADSRRSLMDQGNFPLIKLIKIEDSDTYICEVEDQ-----KEEV 118
Db 3633 -----ENGSHTLVTKDNNADFGKTCQATNKAKGDETVGLKIPKYSFEKQTAEEV 3684

```

```

QY 119 QLVGLTANSPDTHLLOGSLTLTLESPGSSPSVQCRSPRGKNIQGGKTLSSQLELOD 178
Db 3685 KLF--IEPLKTPAVBGDTVLBECKANKESHPOKF-----PKNQPVVIGQMOLFEVL 3738
QY 179 SGTWTCTVLQNGKVEFKIDIVLAFQKASSIVYKKEGQVEFSPPLAFTVEKLTGSGEL 238
Db 3739 DNINIKLTI-QNAKK-----EDVG-----AVRCBAVVAWAGK- 3767
QY 239 WQAARASSSKSWITFDLKNKEVSVKRVTDPKLQMGKLPPLHT-----LPQALPOYAS 294
Db 3768 -----ANTNADKTIQPAKVEHVTDESQLEBIGQFETV 3802
QY 295 GNLTALTEAKTGKHAQEVNLVVMRATOLQK---LTCEVWGPSTPMLSLTENKAKV 351
Db 3803 GD-TASSKTDIORGAPFVPELLRSGCTVTEKQAILKCYKGRPRKIKWT--KEGEVEM 3859
QY 352 SKREKPVW-----VLNPAQMQC-LISDSQVLLSNIKVLPTMSTPVPCA 398
Db 3860 SARVAEHHKDDGTLTLTFDNTVQADAGEYRCBAENEYGSAMTEGPITVLEGAPKIDGA 3919
QY 399 PEPKCDKHTHCPBELLGGSVFLPPPKDITMISRTPEVTCVVVDVSHEDPEVKFNWYV 458
Db 3920 PD-----FLQPVKPAV-----TVGETVAVLEGKISGKRPVYKWK- 3955
QY 459 DQVEVHNAKTKPREQVNSTYRVAVSLTVLHQDWLNGKCKKCVSKALPAPIEKTISK 518
Db 3956 NEBELKPSRVAVIENLDGCTQ---LVTNAKLDDMDYRCBAENEPDWDVTLT-- 4009
QY 519 KQAPREPOVYTLPSRDELTKNQVSLT-----CLVKGFPDIAVWESNQPENNYKT 572
Db 4010 ---VKEP-AQVAPGPFKELSAIQVKETETAKRECKVSGTKPD---VKMFKDGPLKEDXR 4062
QY 573 TPPLVDGSRFLYKSLTVYDKSRM-QQGVNFCSSVMHEA 610
Db 4063 VHESTSDGT---QLVIEDSKTDDQN-VRIEVSND 4096

RESULT 64
LAC2_RAT STANDARD; PRT; 104 AA.
AC P20767;
DT 01-FEB-1991 (rel. 17, Created)
DT 01-FEB-1991 (rel. 17, Last sequence update)
DT 10-OCT-2003 (rel. 42, Last annotation update)
DE Ig lambda-2 chain C region.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87305594; PubMed=3114047;
RA Steen M.L., Hellman U., Petersson U.
RT "The immunoglobulin lambda locus in rat consists of two C lambda
RT genes and a single V lambda gene."
RL Gene 55:75-84(1987).
CC -!- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M2521; AAA1420.1; ALT_INIT.
CC HSSP; P01842; 2MCG.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003597; Ig_C1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig1.
DR SMART; SM00407; IgC1.1.

```



```

FT  VARIANT      302      302      F -> L.
FT  VARIANT      316      316      F -> L.
FT  VARIANT      337      337      G -> R.
FT  VARIANT      367      367      S -> N.
FT  VARIANT      422      422      Q -> L.
FT  VARIANT      429      429      I -> F.
FT  VARIANT      433      433      D -> E.
SQ  SEQUENCE      506 AA; 55093 MM; 6B7E310677FC9CB CRC64;

Query Match      5.0%; Score 171, DB 1; Length 506;
Best Local Similarity 24.4%; Pred. No. 0.0011;
Matches 65; Conservative 40; Mismatches 115; Indels 46; Gaps 9;

QY  400 EPKSCDKTH-----TCPELGGPSVFLPPPKDKDLMISRTPEGVCVVDVSH--EDPE 451
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  133 EFKSGEPTLTVSAKSPPLSGPTV-----RATPEQVNTFTCTSHGSPRN 179
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY  452 VKENMYDVGEVHNAAKTPREEOYNSTRVSVLTVLHODMLNGKEYKCKVSKALPA-- 509
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB  180 ISLKMFKNGNELSASQTSVPDENNNYSINSTTKVLLATGDVHSQVIGCAVAVTLQGGP 239
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY  510 PIKRTISKAKGQPREPOVYTLPPSRDELTL-----KNQVSLTCLVKGFPSPDIAVEWESNG 564
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB  240 PL-----RGTANLSETTIVPPTL-EITGSPSAGNQVNTCCQVNFYFPHQLQTLWLENG 291
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY  565 QPENNYKTPPVLDSDGSFFLYSKLTVDKSRMGGQNVFSGSVNHEALHNHYTKSLSLSP 624
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB  292 NMSRTEASVFEVKDGTFTQTMFLVNSSAHRBAVAVLTCQVHDC-----OPAVSKN 345
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY  625 GLQIDETCAEADQDEL----DGLMTT 646
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB  346 TLEVSAPKQDQDTGQTFGRPNDSNMTS 371
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 66
CEAS HUMAN
ID CEAS_HUMAN STANDARD; PRT; 702 AA.
AC P06731;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Carcinoembryonic antigen-related cell adhesion molecule 5 precursor
DE Carcinoembryonic antigen (CEA) (Meconium antigen 100) (CD666
DE antigen)
GN CEACAM5 OR CEA.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90258861; PubMed=2342461;
RA Schrewe H., Thompson J., Bona M., Hefta L.J.F., Maruya A.,
RA Hasegawa M., Shively J.E., von Kleist S., Zimmermann W.;
RT "Cloning of the complete gene for carcinoembryonic antigen: analysis
RT of its promoter indicates a region conveying cell type-specific
RT expression.";
RL Mol. Cell. Biol. 10:2738-2748(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=88038876; PubMed=3670312;
RA Beaubien N., Benchinol S., Cournoyer D., Fuks A., Stanners C.P.;
RT "Isolation and characterization of full-length functional cDNA clones
RT for human carcinoembryonic antigen.";
RL Mol. Cell. Biol. 7:3221-3230(1987).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=89122014; PubMed=3220478;
RA Barnett T., Goebel S.J., Notdurft M.A., Elting J.J.;
RT "Carcinoembryonic antigen family: characterization of cDNAs coding
RT for CEA and CEA and suggestion of nonrandom sequence variation in
RT their conserved loop-domains.";
RL Genomics 3:59-66(1988).

```

```

RN [4]
RX SEQUENCE OF 5-702 FROM N.A.
RX MEDLINE=87128144; PubMed=3814146;
RA Oikawa S., Nakazato H., Kobeki G.;
RT "Primary structure of human carcinoembryonic antigen (CEA) deduced
RT from cDNA sequence.";
RL Biochem. Biophys. Res. Commun. 142:511-518(1987).
RN [5]
RP SEQUENCE OF 331-702 FROM N.A.
RX MEDLINE=87204247; PubMed=3033671;
RA Zimmermann W., Ortlieb B., Friedrich R., von Kleist S.;
RT "Isolation and characterization of cDNA clones encoding the human
RT carcinoembryonic antigen reveal a highly conserved repeating
RT structure.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:2960-2964(1987).
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -1- TISSUE SPECIFICITY: Found in adenocarcinomas of endoderially
CC derived digestive system epithelium and fetal colon.
CC -1- PTM: COMPLEX IMMUNOREACTIVE GLYCOPROTEIN WITH A MW OF 180 KDa
CC COMPRISING 608 CARBOHYDRATE.
CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily. CEA family.
CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD666 entry;
CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd666.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.ch).
CC -----
DR EMBL: M17303; AAB59513.1; -;
DR EMBL: M59262; AAA62835.1; ALT_SEQ.
DR EMBL: M59255; AAA62835.1; JOINED.
DR EMBL: M59256; AAA62835.1; JOINED.
DR EMBL: M59257; AAA62835.1; JOINED.
DR EMBL: M59258; AAA62835.1; JOINED.
DR EMBL: M59259; AAA62835.1; JOINED.
DR EMBL: M59260; AAA62835.1; JOINED.
DR EMBL: M59261; AAA62835.1; JOINED.
DR EMBL: M59709; -; NOT_ANNOTATED_CDS.
DR EMBL: M59710; -; NOT_ANNOTATED_CDS.
DR EMBL: M29540; AAA51957.1; -;
DR EMBL: X16455; CAA3474.1; -;
DR EMBL: M15042; AAA51963.1; -;
DR EMBL: M16234; AAA51972.1; -;
DR PIR: A36319; A36319.
DR PDB: 1B07; 04-JUL-00.
DR Genew; HGNC:1817; CEACAM5.
DR MIM; 114890; -;
DR GO; GO:0005887; C:integral to plasma membrane; TAS.
DR InterPro; IPR007110; IG-like.
DR Pfam; PF00047; IG; 6.
DR PROSITE; PS50835; IG_LIKE; 6.
KW Immunoglobulin domain, Glycoprotein, Lipoprotein, GPI-anchor;
KW Membrane; Signal; Repeat; 3D-structure.
FT SIGNAL 1 34
FT CHAIN 35 685
FT PROPEP 686 702
FT DOMAIN 35 144
FT DOMAIN 146 237
FT DOMAIN 238 322
FT DOMAIN 324 415
FT DOMAIN 416 498
FT DOMAIN 502 593
FT DOMAIN 594 677
FT LIPID 685 685
FT CARBOHYD 104 104
FT CARBOHYD 115 115
FT CARBOHYD 152 152

```

```

FT CARBOHYD 182 182 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 197 197 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 204 204 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 208 208 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 246 246 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 256 256 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 274 274 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 288 288 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 292 292 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 309 309 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 351 351 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 360 360 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 375 375 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 432 432 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 466 466 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 480 480 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 508 508 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 529 529 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 553 553 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 560 560 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 580 580 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 612 612 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 650 650 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 665 665 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CONFLICT 320 320 MISSING (IN REF. 4)
SO SEQUENCE 702 AA; 76795 MW; 6299AE26C0DBB5C CRC64;

```

Query Match 4.8%; Score 165; DB 1; Length 702;
 Best Local Similarity 17.6%; Pred. No. 0.0039;
 Matches 118; Conservative 109; Mismatches 255; Indels 190; Gaps 31;

```

QY 20 PATQGNKVVVGKGDVTELTCTASQKSIQFHWKNSQIKILNQSGFLTKGSKLND 79
DB 147 PSISSNNSKPEDEK-DAVAFCEPTQDATTLMWVNSLPV----- 187
QY 80 ADSRRSLWDQGNFPLIKLKIEDSDYICEVED-----QKEEYQV-LVFGLTANS- 129
DB 188 --SRLQSGNNGNRITLTFNVRNDPTASVKCEQNQVNSARSDSYILANLYGPDAPPTISPL 245
QY 130 DTHLQGSLLTTLSPGSSPSVQCR-SPRKNIGSGKTLVSQLEIDSGTCTVYLQ 188
DB 246 NTSYRSGENLMSCHA--ASNPPAOYSFVNGTFOQSTQELFIRPITVNNSGSYTCQAHN 303
QY 189 NQKNV-EFKIDIVLAFQKASSIYKKGEQVESFPIATFVE-KLNGSGSLMQAERAS 246
DB 304 SDTGUNRTTVTITVYAEPPKPTSSNNSNPEDEDAVALTCEPIONITVYLMV----- 357
QY 247 SSKSWITFDLKNKEVSVKRVTQDPKLQMGKLLPHLLTPQALPOYAGSGNLTLLAEATG 306
DB 358 -----VANGSLPV-----SPRLQ-----SNDNTLTLLSVTR 385
QY 307 KLHGVNLVVRATQLOKNTLCE-----VWGPTSP-----KLMLSLKL 344
DB 386 -----NDVGPEYEGCIQNELSVDSHSDPVILNVLVGPDPPTISPTYYRPGVNLISCHA 439
QY 345 ENK-----EAKVSKREKRPVWVLN---PEAGMQLDSDSQOVLESIKYLPMTST 392
DB 440 ASNPPAOYSWMLIDENIQHTQELFISNTEKNSGLYTQANNSSAGHSRTTVKTI-TVSA 498
QY 393 PVPCPAPRPSKCDKHTHCPRELLGGPSVFLPPPKPDOTLMSRTEPVTCVVAVDSHEBDEV 452
DB 499 ELPRPSISSNNSK-----FVEDKDAVAFCEBEA-----ONT 530
QY 453 KFNWYVDGVEVNAKTKPREQVNSTYRVSVLTVLHODWLNKGEYKCKVSNKALPAPIE 512
DB 531 TYLWVWNG---QSLPVSPRLQLSNGN-RTLTFLFVTRND--ABAAYCGIGN----- 575
QY 513 KTISAKQGPREFQV-----YTLPSRDELTKNOVSLTCLVKGFTYPSDIAVWESNGQ 565
DB 576 -SVSANRSDPVTLVDLVPDTPITISPPSSYLSGANLWSC-HSASNSP-QYSWRINGI 632
QY 566 PENNYKTPPVLDSDGSFFLYSKLTVDSKRWQGVFSCVMHEAL--HNHYTKSLSD-- 622

```

```

DB 633 PQOHTQV-----LPIAKTIPNN-----GTACFVSNLATGRNNSIVKSYTVSA 676
QY 623 ---SPGLQDET 631
DB 677 SGTSFGLSAGAT 688

```

RESULT 67

ID LAC3_MOUSE

STANDARD; PRT; 104 AA.

AC P01845;

21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986

(Rel. 01, Last sequence update)

DE 10-OCT-2003

(Rel. 42, Last annotation update)

OS Ig lambda-3 chain C region.

OC Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_Taxid=10090;

RN [1]

RP MEDLINE=83014953; PubMed=6812053;

RA Seising E., Miller J., Wilson R., Scorb U.;

RT "Evolution of mouse immunoglobulin lambda genes."

RL Proc. Natl. Acad. Sci. U.S.A. 79:4681-4685(1982).

RN [2]

RP SEQUENCE OF 1-18 (MYELOMA PROTEIN CBPC-49).

RA Breyer R.M., Sauer R.T., Eisen H.N.;

RT "The variable region of mouse lambda-3 chains."

RL ICN UCLIA Symp. Mol. Cell. Biol. 20:105-110(1981).

RN [3]

RP SEQUENCE OF 10-104 (MYELOMA PROTEIN CBPC-49 AND MONOCLONAL ANTIBODY

RP 8-47).

RX MEDLINE=8123782; PubMed=6165998;

RA Azuma T., Steiner L.A., Eisen H.N.;

RT "Identification of a third type of lambda light chain in mouse

RL immunoglobulins."

CC Proc. Natl. Acad. Sci. U.S.A. 78:569-573(1981).

CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>CC or send an email to license@isb-sib.ch).

CC -----

DB EMBL; J00585; AB59670.1; -

DB PIR; B93922; LMSG.

DB HSSP; P01842; 2MCG.

DB InterPro; IPR007110; Ig-like.

DB InterPro; IPR003597; Ig cl.

DB InterPro; IPR003006; Ig_MHC.

DB Pfam; PF00047; Ig_1.

DB SMART; SM00407; Igcl; 1.

DB PROSITE; PS00835; Ig_LIKE; 1.

DB PROSITE; PS00290; Ig_MHC; 1.

KM Immunoglobulin domain; Immunoglobulin C region.

FT NON_TER 1 1

FT DOMAIN 6 99

FT DISULFID 27 85

FT DISULFID 103 103

SO SEQUENCE 104 AA; 11371 MW; 83CEBCD4AA348EF1 CRC64;

Query Match 4.8%; Score 163; DB 1; Length 104;
 Best Local Similarity 34.3%; Pred. No. 0.0043;
 Matches 35; Conservative 16; Mismatches 47; Indels 4; Gaps 2;

```

QY 521 QPRE-POVYTLPPSRDELTKNOVSLTCLVKGFTYPSDIAVWESNGQENNYKTPPVLD 579
DB 1 QPKSTPLTWMPSPRELQENKATLVCLISNFSBPGVTVMKANGFTITGQVDSNPTKE 60

```

QY 580 DGSFPLYSKLTVDKSRWQGNVFSQSVMEALHNHTQKSL 621
DB 61 DNRYMASSFLLHITSDDWRSHNSFTQVTHE---GDTVEKSL 99

RESULT 68
ID IL1L1_HUMAN STANDARD; PRT; 213 AA.
AC P15814;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Immunoglobulin lambda-like polypeptide 1 precursor (Immunoglobulin-
DE related 14.1 protein) (Immunoglobulin omega polypeptide) (lambda 5)
DE (CD179b antigen).
GN IG1L1 OR IG1L.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89315835; PubMed=2501791;
RA Hollis G.F., Evans R.J., Stafford-Hollis J.M., Korsmeyer S.J.,
RA McKearn J.P.;
RT "Immunoglobulin lambda light-chain-related genes 14.1 and 16.1 are
RT expressed in pre-B cells and may encode the human immunoglobulin
RT omega light-chain protein.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:5552-5556(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymphoid;
RX MEDLINE=91108327; PubMed=1703205;
RA Evans R.J., Hollis G.F.;
RT "Genomic structure of the human Ig lambda 1 gene suggests that it may
RT be expressed as an Ig lambda 14.1-like protein or as a canonical B
RT cell Ig lambda light chain: implications for Ig lambda gene
RT evolution.";
RL J. Exp. Med. 173:305-311(1991).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,
RA Diatchenko L., Mansura K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uebin T.B., Tomihata S., Carninci P., Prange C.,
RA Roha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butcherfield Y.S.N., Krzywinski M.I., Skalek U., Smallus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -1- SUBUNIT: Associates non-covalently with VREB1.
CC -1- TISSUE SPECIFICITY: EXPRESSED ONLY IN PRE-B-CELLS AND A SPECIAL
CC B-CELL LINE (WHICH IS SURFACE IG NEGATIVE).
CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily.
CC -1- DATABASE: NMBE=PROV; NOTE=PROV 1:64-67(2000);
CC WWW="http://www.ncbi.nlm.nih.gov/prov/guide/596419174.g.htm".
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M27749; AAA36100.1; -
DR EMBL; M34513; AAA36096.1; -
DR EMBL; M34511; AAA36096.1; JOINED.
DR EMBL; M34512; AAA36096.1; JOINED.
DR EMBL; BC012293; AAA12293.1; -
DR PIR; A33911; A33911.
DR HSSP; P01842; 7FAR.
DR Genew; HGNC:5870; IG1L1.
DR MIM; 146770; -
DR GO; GO:0016020; C:membrane; NAS.
DR GO; GO:0003793; F:defense/immunity protein activity; NAS.
DR GO; GO:0006955; P:immune response; NAS.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003587; Ig-CL.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig1; 1.
DR SMART; SM00407; IG1; 1.
DR PROSITE; PS00835; IG_LIKE; 1.
DR PROSITE; PS00280; IG_MHC; 1.
KW Antigen; Signal; Immunoglobulin domain.
FT SIGNAL 1
FT CHAIN 37
FT CHAIN 38
FT DOMAIN 97
FT DOMAIN 108
FT DOMAIN 109
FT DOMAIN 213
SQ SEQUENCE 213 AA; 22963 MW; 9133A7742B943C79 CRC64;
Query Match 4.8%; Score 162.5; DB 1; Length 213;
Best Local Similarity 30.5%; Pred. No. 0.0012;
Matches 36; Conservative 21; Mismatches 44; Indels 17; Gaps 3;
QY 521 QPR-EPQVTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQP-ENNYKTPPVLD 578
DB 109 QPKATPSVTLFPSSSEELQANKATLVCLMNDYFPGILTVTKADGTPITQGVEMTPPSKQ 168
QY 579 DGSFPLYSKLTVDKSRWQGNVFSQSVMEALHNHTQKSL 621
DB 169 SNNKTAASSYLSLTPEQWRSRRSRYSQVWHE-----GSTVEKTVAPAE 211

RESULT 69
ID NE01_MOUSE STANDARD; PRT; 1493 AA.
AC P97788;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Neogenin precursor.
DE Neogenin precursor.
GN NE01 OR NGN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC TISSUE=Brain;
RX MEDLINE=97407661; PubMed=9264410;
RA Keeling S.L., Gad J.M., Cooper H.M.;
RT "Mouse neogenin, a DCC-like molecule, has four splice variants and is
RT expressed widely in the adult mouse and during embryogenesis.";
RL Oncogene 15:691-700(1997).
CC -1- FUNCTION: May be involved as a regulatory protein in the
CC transduction of undifferentiated proliferating cells to their
CC differentiated state. May also function as a cell adhesion
CC molecule in a broad spectrum of embryonic and adult tissues.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.

CC	-1	ALTERNATIVE PRODUCTS:	
CC		Event=Alternative splicing; Named isoforms=5;	
CC		Comment=Additional isoforms seem to exist;	
CC		Name=1;	
CC		Isoid=p97798-1; Sequence=Displayed;	
CC		Name=2;	
CC		Isoid=p97798-2; Sequence=VSP_002594;	
CC		Name=3;	
CC		Isoid=p97798-3; Sequence=VSP_002595;	
CC		Note=Expression developmentally regulated;	
CC		Name=4;	
CC		Isoid=p97798-4; Sequence=VSP_002596;	
CC		Note=Expression developmentally regulated;	
CC		Name=5;	
CC		Isoid=p97798-5; Sequence=VSP_002597;	
CC		Note=Expression developmentally regulated;	
CC	-1	TISSUE SPECIFICITY: Widely expressed.	
CC	-1	DEVELOPMENTAL STAGE: EXPRESSED UNICITOUSLY THROUGHOUT THE MID TO LATE STAGES OF GESTATION AND IN ADULT TISSUES. STRONG EXPRESSION IS OBSERVED IN THE VENTRAL REGION OF THE VENTRICULAR ZONE OF THE E15.5 MOUSE NEURAL TUBE, AS WELL AS IN THE VENTRICULAR ZONES OF THE MESENCEPHALON AND RHOMBENCEPHALON. ISOFORMS 3 AND 4 ARE EXPRESSED AT HIGHER LEVEL COMPARED TO OTHER ISOFORMS BETWEEN E11.5 AND E6.5.	
CC	-1	SIMILARITY: Belongs to the immunoglobulin superfamily, DCC family.	
CC	-1	SIMILARITY: Contains 4 immunoglobulin-like C2-type domains.	
CC	-1	SIMILARITY: Contains 6 fibronectin type III domains.	
CC	-1	-----	
CC		This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).	
CC		-----	
DR	EMBL	Y09535; CAAT70727.1; -.	
DR	HSSP	P02751; ITTF.	
DR	MGI	MGI:1097159; Neol.	
DR	Interpro	IPR008957; FN_III-like.	
DR	Interpro	IPR003961; FN_III.	
DR	Interpro	IPR003962; FnIII subd.	
DR	Interpro	IPR007110; Ig-like.	
DR	Interpro	IPR003598; Ig_C2.	
DR	Pfam	PF00047; fn3; 6.	
DR	Pfam	PF00047; Ig; 4.	
DR	PRINTS	PR00014; FNTPR111.	
DR	SMART	SM00060; FN3; 6.	
DR	SMART	SM00408; IGc2; 4.	
DR	PROSITE	PS50835; IG_LIKE; 4.	
KW	Cell adhesion; Repeat; Signal; Transmembrane; Immunoglobulin domain; Glycoprotein; Alternative splicing.		
KW	Glycoprotein; Alternative splicing.		
FT	SIGNAL	1 35	POTENTIAL.
FT	CHAIN	37 1493	NEOGENIN.
FT	DOMAIN	37 1136	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1137 1157	POTENTIAL.
FT	DOMAIN	1158 1493	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	63 158	IG-LIKE C2-TYPE 1.
FT	DOMAIN	163 249	IG-LIKE C2-TYPE 2.
FT	DOMAIN	254 347	IG-LIKE C2-TYPE 3.
FT	DOMAIN	352 437	IG-LIKE C2-TYPE 4.
FT	DOMAIN	467 564	FIBRONECTIN TYPE-III 1.
FT	DOMAIN	567 660	FIBRONECTIN TYPE-III 2.
FT	DOMAIN	661 760	FIBRONECTIN TYPE-III 3.
FT	DOMAIN	766 860	FIBRONECTIN TYPE-III 4.
FT	DOMAIN	861 981	FIBRONECTIN TYPE-III 5.
FT	DOMAIN	982 1083	FIBRONECTIN TYPE-III 6.
FT	DOMAIN	1149 1153	POLY-VAL.
FT	DISULFID	85 140	BY SIMILARITY.
FT	DISULFID	184 232	BY SIMILARITY.
FT	DISULFID	281 331	BY SIMILARITY.
FT	DISULFID	373 421	BY SIMILARITY.
FT	CARBOHYD	84 84	N-LINKED (GLCNAC. . .) (POTENTIAL).

Query Match	Best Local Similarity	4.8%; Score 162.5; DB 1; Length 1493;	21.8%; Pred. No. 0.015;	Matches 159; Conservative 79; Mismatches 282; Indels 211; Gaps 40
FT CARBOHYD	221	221	N-LINKED (GLCNAC . . .)	(POTENTIAL)
FT CARBOHYD	337	337	N-LINKED (GLCNAC . . .)	(POTENTIAL)
FT CARBOHYD	501	501	N-LINKED (GLCNAC . . .)	(POTENTIAL)
FT CARBOHYD	520	520	N-LINKED (GLCNAC . . .)	(POTENTIAL)
FT CARBOHYD	670	670	N-LINKED (GLCNAC . . .)	(POTENTIAL)
FT CARBOHYD	746	746	N-LINKED (GLCNAC . . .)	(POTENTIAL)
FT CARBOHYD	940	940	N-LINKED (GLCNAC . . .)	(POTENTIAL)
FT VASPPLIC	442	461	Missing (In isoform 2).	
FT VASPPLIC	863	878	/FtId=VSP_002594.	
FT VASPPLIC	1086	1096	Missing (In isoform 3).	
FT VASPPLIC	1279	1331	/FtId=VSP_002595.	
FT VASPPLIC	1279	1331	Missing (In isoform 4).	
FT VASPPLIC	1279	1331	/FtId=VSP_002596.	
FT VASPPLIC	1279	1331	Missing (In isoform 5).	
FT VASPPLIC	1279	1331	/FtId=VSP_002597.	
FT VASPPLIC	1279	1331	Missing (In isoform 6).	
FT VASPPLIC	1279	1331	Missing (In isoform 7).	
FT VASPPLIC	1279	1331	Missing (In isoform 8).	
FT VASPPLIC	1279	1331	Missing (In isoform 9).	
FT VASPPLIC	1279	1331	Missing (In isoform 10).	
FT VASPPLIC	1279	1331	Missing (In isoform 11).	
FT VASPPLIC	1279	1331	Missing (In isoform 12).	
FT VASPPLIC	1279	1331	Missing (In isoform 13).	
FT VASPPLIC	1279	1331	Missing (In isoform 14).	
FT VASPPLIC	1279	1331	Missing (In isoform 15).	
FT VASPPLIC	1279	1331	Missing (In isoform 16).	
FT VASPPLIC	1279	1331	Missing (In isoform 17).	
FT VASPPLIC	1279	1331	Missing (In isoform 18).	
FT VASPPLIC	1279	1331	Missing (In isoform 19).	
FT VASPPLIC	1279	1331	Missing (In isoform 20).	
FT VASPPLIC	1279	1331	Missing (In isoform 21).	
FT VASPPLIC	1279	1331	Missing (In isoform 22).	
FT VASPPLIC	1279	1331	Missing (In isoform 23).	
FT VASPPLIC	1279	1331	Missing (In isoform 24).	
FT VASPPLIC	1279	1331	Missing (In isoform 25).	
FT VASPPLIC	1279	1331	Missing (In isoform 26).	
FT VASPPLIC	1279	1331	Missing (In isoform 27).	
FT VASPPLIC	1279	1331	Missing (In isoform 28).	
FT VASPPLIC	1279	1331	Missing (In isoform 29).	
FT VASPPLIC	1279	1331	Missing (In isoform 30).	
FT VASPPLIC	1279	1331	Missing (In isoform 31).	
FT VASPPLIC	1279	1331	Missing (In isoform 32).	
FT VASPPLIC	1279	1331	Missing (In isoform 33).	
FT VASPPLIC	1279	1331	Missing (In isoform 34).	
FT VASPPLIC	1279	1331	Missing (In isoform 35).	
FT VASPPLIC	1279	1331	Missing (In isoform 36).	
FT VASPPLIC	1279	1331	Missing (In isoform 37).	
FT VASPPLIC	1279	1331	Missing (In isoform 38).	
FT VASPPLIC	1279	1331	Missing (In isoform 39).	
FT VASPPLIC	1279	1331	Missing (In isoform 40).	
FT VASPPLIC	1279	1331	Missing (In isoform 41).	
FT VASPPLIC	1279	1331	Missing (In isoform 42).	
FT VASPPLIC	1279	1331	Missing (In isoform 43).	
FT VASPPLIC	1279	1331	Missing (In isoform 44).	
FT VASPPLIC	1279	1331	Missing (In isoform 45).	
FT VASPPLIC	1279	1331	Missing (In isoform 46).	
FT VASPPLIC	1279	1331	Missing (In isoform 47).	
FT VASPPLIC	1279	1331	Missing (In isoform 48).	
FT VASPPLIC	1279	1331	Missing (In isoform 49).	
FT VASPPLIC	1279	1331	Missing (In isoform 50).	
FT VASPPLIC	1279	1331	Missing (In isoform 51).	
FT VASPPLIC	1279	1331	Missing (In isoform 52).	
FT VASPPLIC	1279	1331	Missing (In isoform 53).	
FT VASPPLIC	1279	1331	Missing (In isoform 54).	
FT VASPPLIC	1279	1331	Missing (In isoform 55).	
FT VASPPLIC	1279	1331	Missing (In isoform 56).	
FT VASPPLIC	1279	1331	Missing (In isoform 57).	
FT VASPPLIC	1279	1331	Missing (In isoform 58).	
FT VASPPLIC	1279	1331	Missing (In isoform 59).	
FT VASPPLIC	1279	1331	Missing (In isoform 60).	
FT VASPPLIC	1279	1331	Missing (In isoform 61).	
FT VASPPLIC	1279	1331	Missing (In isoform 62).	
FT VASPPLIC	1279	1331	Missing (In isoform 63).	
FT VASPPLIC	1279	1331	Missing (In isoform 64).	
FT VASPPLIC	1279	1331	Missing (In isoform 65).	
FT VASPPLIC	1279	1331	Missing (In isoform 66).	
FT VASPPLIC	1279	1331	Missing (In isoform 67).	
FT VASPPLIC	1279	1331	Missing (In isoform 68).	
FT VASPPLIC	1279	1331		

RESULT 70
 ID KMLS_CHICK STANDARD; PRT; 1906 AA.
 AC P11799; P19038;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Myosin light chain kinase, smooth muscle and non-muscle isozymes
 DE (EC 2.7.1.117) (MCKK) [Contains: Telokin].
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 NC NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM MLCK-210).
 RX MEDLINE=96033976; PubMed=7589469;
 RA Watterson D.M., Collinge M., Lukas T.J., van Eldik L.J., Birukov K.G.,
 RA Stepanova O.V., Shtrinsky V.P.,
 RT "Multiple gene products are produced from a novel protein kinase
 RT transcription region.";
 RL FEBS Lett. 373:217-220(1995).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM MLCK-108).
 RX MEDLINE=90192792; PubMed=2315320;
 RA Olson N.J., Pearson R.B., Needleman D.S., Hurwitz M.J., Kemp B.E.,
 RA Means A.R.,
 RT "Regulatory and structural motifs of chicken gizzard myosin light
 RT chain kinase.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:2284-2288(1990).
 RN [3]
 RP SEQUENCE OF 649-1906 FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Fibroblast;
 RX MEDLINE=90361738; PubMed=2202734;
 RA Shoemaker M.O., Lau W., Shattuck R.L., Kwiatkowski A.P.,
 RA Mastrisani P.E., Guerra-Santos L., Wilson E., Lukas T.J.,
 RA van Eldik L.J., Watterson D.M.,
 RT "Use of DNA sequence and mutant analyses and antisense
 RT oligodeoxynucleotides to examine the molecular basis of nonmuscle
 RT myosin light chain kinase autoinhibition, calmodulin recognition, and
 RT activity.";
 RL J. Cell Biol. 111:1107-1125(1990).
 RN [4]
 RP SEQUENCE OF 1259-1906 FROM N.A.
 RC TISSUE=Gizzard;
 RX MEDLINE=9157587; PubMed=3030394;
 RA Gueriere V. Jr., Russo M.A., Olson N.J., Putkey J.A., Means A.R.,
 RT "Domain organization of chicken gizzard myosin light chain kinase
 RT deduced from a cloned cDNA.";
 RL Biochemistry 25:8372-8381(1986).
 RN [5]
 RP SEQUENCE OF 1750-1906 FROM N.A. (ISOFORM TELOKIN).
 RC TISSUE=Gizzard;
 RX MEDLINE=93073972; PubMed=1444462;
 RA Yoshikai S., Ikebe M.,
 RT "Molecular cloning of the chicken gizzard telokin gene and cDNA.";
 RL Arch. Biochem. Biophys. 299:242-247(1992).
 RN [6]
 RP SEQUENCE OF 1750-1906 FROM N.A. (ISOFORM TELOKIN).
 RX MEDLINE=92236611; PubMed=1373815;
 RA Collinge M., Mastrisani P.E., Zimmer W.E., Shattuck R.L., Lukas T.J.,
 RA van Eldik L.J., Watterson D.M.,
 RT "Structure and expression of a calcium-binding protein gene contained
 RT within a calmodulin-regulated protein kinase gene.";
 RL Mol. Cell. Biol. 12:2359-2371(1992).
 CC -1- FUNCTION: Phosphorylates a specific serine in the N-terminus of a
 CC myosin light chain, which leads to the formation calmodulin/MLCK
 CC signal transduction complexes which allow selective transduction
 CC of calcium signals.
 CC -1- CATALYTIC ACTIVITY: ATP + [myosin light-chain] = ADP + [myosin
 CC light-chain] phosphate.
 CC -1- ALTERNATIVE PRODUCTS:

CC Event-Alternative Initiation; MLCK-210/Non-muscle
 CC Comment-At least 3 isoforms: MLCK-210/Non-muscle
 CC MLCK-108/Smooth-muscle and Telokin, are produced by alternative
 CC Initiation;
 CC -1- TISSUE SPECIFICITY: ISOFORM TELOKIN IS EXPRESSED IN GIZZARD,
 CC HEART, LUNG, INTESTINE, AND SKELETAL MUSCLE ALTHOUGH THE LEVELS OF
 CC THE EXPRESSION IN THE LATTER WERE MUCH LESS THAN THAT IN THE
 CC GIZZARD.
 CC -1- DOMAIN: TELOKIN BINDS CALMODULIN.
 CC -1- SIMILARITY: Belongs to the Ser/Thr family of protein kinases.
 CC -1- SIMILARITY: Contains 9 immunoglobulin-like C2-type domains.
 CC -1- SIMILARITY: Contains 1 fibronectin type III domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X52876; CAA37056.1; -
 CC EMBL; X52876; CAA37057.1; -
 CC EMBL; X52876; CAA37058.1; -
 CC EMBL; M31048; AAA49069.1; -
 CC EMBL; M14953; AAA69964.1; -
 CC EMBL; M96655; AAA90083.1; -
 CC EMBL; M88283; AAA48647.1; -
 CC EMBL; M88284; AAB53768.1; -
 CC PIR; S68235; S68235.
 CC PDB; 1CDL; 31-AUG-94.
 CC PDB; 1VRK; 27-APR-99.
 CC InterPro; IPR008957; FN_III-like.
 CC InterPro; IPR003961; FN_III.
 CC InterPro; IPR007110; IG-like.
 CC InterPro; IPR003598; IG_C2.
 CC InterPro; IPR000719; Prot_kinase.
 CC InterPro; IPR002290; Ser_thr_pkin_AS.
 CC InterPro; IPR000411; fn3; 1.
 CC Pfam; PF00047; Ig; 9.
 CC Pfam; PF00069; pkinase; 1.
 CC ProDom; PD000001; Prot_kinase; 1.
 CC SMART; SM00060; FN3; 1.
 CC SMART; SM00408; IGC2; 8.
 CC SMART; SM00220; S_TKC; 1.
 CC PROSITE; PS50835; IG_LIKE; 9.
 CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 CC PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 CC PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 CC Transferrase; Serine/threonine-protein kinase; Calmodulin-binding;
 CC ATP-binding; Phosphorylation; Immunoglobulin domain; Repeat;
 CC Alternative Initiation; 3D-structure.
 CC CHAIN 1 1906
 CC FT MYOSIN LIGHT CHAIN KINASE, ISOFORM MLCK-
 CC FT 210.
 CC CHAIN 935 1906
 CC FT MYOSIN LIGHT CHAIN KINASE, ISOFORM MLCK-
 CC FT 108.
 CC CHAIN 1750 1906
 CC FT MYOSIN LIGHT CHAIN KINASE, ISOFORM
 CC FT TELOKIN.
 CC FT INIT MET 935 935
 CC FT FOR ISOFORM MLCK-108.
 CC FT INIT MET 1750 1750
 CC FT FOR ISOFORM TELOKIN.
 CC FT DOMAIN 156 117
 CC FT IG-LIKE C2-TYPE 1.
 CC FT DOMAIN 156 244
 CC FT IG-LIKE C2-TYPE 2.
 CC FT DOMAIN 429 517
 CC FT IG-LIKE C2-TYPE 3.
 CC FT DOMAIN 521 613
 CC FT IG-LIKE C2-TYPE 4.
 CC FT DOMAIN 637 725
 CC FT IG-LIKE C2-TYPE 5.
 CC FT DOMAIN 735 830
 CC FT IG-LIKE C2-TYPE 6.
 CC FT DOMAIN 1084 1172
 CC FT IG-LIKE C2-TYPE 7.
 CC FT DOMAIN 1225 1313
 CC FT IG-LIKE C2-TYPE 8.
 CC FT DOMAIN 1330 1400
 CC FT FIBRONECTIN TYPE-III.
 CC FT DOMAIN 1453 1708
 CC FT PROTEIN KINASE.
 CC FT DOMAIN 1794 1885
 CC FT IG-LIKE C2-TYPE 9.
 CC NP_BIND 1459 1467
 CC ATP (BY SIMILARITY).

```

FT BINDING 1482 1482 ATP (BY SIMILARITY)
FT ACT SITE 1574 1574 BY SIMILARITY
FT DOMAIN 1716 1728 CALMODULIN AUTOINHIBITION (AM13) REGION
FT 1730 1749 (POTENTIAL)
FT 1730 1749 CALMODULIN RECOGNITION (RS20) REGION
FT 1730 1749 (POTENTIAL)
FT DOMAIN 1317 1364 MOTIF IA.
FT 1385 1402 MOTIF IB.
FT 1833 1833 4 X REPEATS, MOTIF IIA.
FT 660 676 IIA-1.
FT 758 774 IIA-2.
FT 1107 1123 IIA-3.
FT 1817 1833 IIA-4.
FT 693 1866 5 X REPEATS, MOTIF IIB.
FT 791 807 IIB-1.
FT 1140 1156 IIB-2.
FT 1281 1297 IIB-3.
FT 1851 1866 IIB-4.
FT 970 1226 IIB-5.
FT 970 987 4 X REPEATS, MOTIF III.
FT 999 1016 IIB-1.
FT 1061 1078 IIB-2.
FT 1209 1226 IIB-3.
FT 1700 1763 CALMODULIN-BINDING.
FT 1896 1906 POLY-GLU.
FT 1748 1748 PHOSPHORYLATION.
FT MOD_RES 1762 1762 PHOSPHORYLATION.
FT MOD_RES 1439 1439 R -> Q (IN REF. 4).
FT CONFLICT 1439 1439
SQ SEQUENCE 1906 AA; 210445 MW; AD7D8A3B69EE3163 CRC64;

```

Query March 4.7%; Score 162; DB 1; Length 1906;

Best Local Similarity 21.3%; Pred. No. 0.022; Matches 150; Conservative 89; Mismatches 256; Indels 208; Gaps 34;

```

QY 89 QGNPFLIKNLKIDSDTYICEVED---QKEEVQLVFGGLTANSDDLLOGSLTTLLE 144
DB 81 RGISLIVKIGVEGDSKGTCEANDGVRQVTLVEG---NS---LKKYSLPSSAK 133
QY 145 SPPG--SSPSVQCR-----SPRKNIQGGKT----- 168
DB 134 TPGRLSVPVVHRPPIWSESPKATKPNRVVREGQTRFSCKITGRPOQVWTMGD 193
QY 169 -----LSVQLELQDSTGWTCTVQLQNKYKFEKDIYV----- 201
DB 194 IHLQNERFMFEKGTGYLETIQNVQLADAGIYCTVNSGKASVSALFTVQGPDKTDT 253
QY 202 -----LA-----FQKASSIVYKKEGEQVEFSPLAFTVEKLTGSGELMW 240
DB 254 HAQPLCMPPKPTTLATKAIEKSDFOATNSGIAKEIKSTREL-MVETKDLISAKETFY 312
QY 241 CAERASSSSKSWITTDLKNKEVSVKRVTDPKLQNGKKLPLHLTPQALPYAGSGNLTLA 300
DB 313 TSRAKQKQKQONQONANAVPQESRGTGPQLKTSSTITLQAVKAPBEKAPQTFPI 372
QY 301 LEATKGLKQGVNLVWMAATQLOKLTCEVWGPPSPKMLSLKLEN-----KEAKVSKRE 355
DB 373 ROAEDRK--RTVQPLMTTITQENPSLT---GVVSPR---SRETEENRAGVKSVEEKRE 423
QY 356 KPVWLVNLEAGMOCCLSDSGO-VLLESNIKVLPT-----WSTVPCCAPAPKSC---DK 406
DB 424 -PLGIPOFESRPSLEASBEGQELKFKSKVSGKRPDVEMKEGVPIITGSGIQTIEEDG 482
QY 407 TH-----TCPPELLGGPSVLFPPPKPDT--LMISRTPEVT-----CVVVDV 445
DB 483 THCLMLKKAICLGDGSGYSKAFNPGRGTSMLLTVKAPKEVAVPCFSSVLKQCTV--- 539
QY 446 SHDEPVKFMVYVGVVHNAKTKRER--QY-NSTYV--VSULTVYLHODPLNKE--YK 499
DB 540 -SEGQDFVLQCYGVGVVPBEITWLLNEQPIQYASTFEAGAKLTV--QDALPBDGITYT 596
QY 500 CKVSNKA-----LPAPIETKISKAGQPREPOVYVLP-----SRDELTKNOVSL 544

```

```

DB 597 CLAENNAGRASCAQVTVKREKSSKKAEGTQAANKLPAPILFKGLTDLKWDGSGVIM 656
QY 545 TCLVKGFPYDIAVEMESNGQPE-----NNYKTPPVLD-----SDGSFFLYSKLTVDK 593
DB 657 T-----VEVSANPCPELITWLNKGEKIQETDHFPEKKGNYSLYIQGVFPEDT 704
QY 594 SRWQGVFSCSVWHEALNNHYTQKSLSPGLQLDFTCAADQ 638
DB 705 GK-----YTCSEANNE-LGETQQTQATL-----TVQEPDQG 732

RESULT 71
LAC2 MOUSE
ID LAC2 MOUSE STANDARD; PRT; 104 AA.
AC P01844;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig lambda-2 chain C region.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=63014953; PubMed=6812053;
RA Seising E., Miller J., Wilson R., Storb U.;
RT "Evolution of mouse immunoglobulin lambda genes.";
RL Proc. Natl. Acad. Sci. U.S.A. 79:4681-4685(1982).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=62274221; PubMed=6287422;
RA Wu G.E., Govindji N., Hozumi N., Murialdo H.;
RT "Nucleotide sequence of a chromosomal rearranged lambda 2
RL immunoglobulin gene of mouse.";
RN [3]
RX MEDLINE=6220143; PubMed=6283385;
RA Botwell A.L.M., Paskind M., Reith M., Imanishi-Kari T., Rajewsky K.,
RL Baltimore D.;
RT "Somatic variants of murine immunoglobulin lambda light chains.";
RL Nature 298:380-382(1982).
[4]
RP SEQUENCE (MOPC 315).
RX MEDLINE=74048693; PubMed=4760498;
RA Dugan E.S., Bradshaw R.A., Simms E.S., Eisen H.N.;
RT "Amino acid sequence of the light chain of a mouse myeloma protein
RL (MOPC-315).";
RN [5]
RL Biochemistry 12:5400-5416(1973).
RP SEQUENCE OF 66-104 (MOPC 315), AND REVISIONS.
RX MEDLINE=81223782; PubMed=6165998;
RA Azuma T., Steiner L.A., Eisen H.N.;
RT "Identification of a third type of lambda light chain in mouse
RL immunoglobulins.";
RL Proc. Natl. Acad. Sci. U.S.A. 78:569-573(1981).
CC -1-SIMILARITY: Contains 1 immunoglobulin-like domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk.)
CC -----
DB EMBL; J00595; AAA39151.1; -
DB F01844; L2MS.
DB HSSP; P01842; 2MCG.
DB InterPro; IPR007110; Ig_1like.
DB InterPro; IPR003597; Ig_c1.
DB InterPro; IPR003006; Ig_MHC.

```

DR Pfam: PF00047; Ig; 1.
 DR SMART: SM00407; IgC1; 1.
 DR PROSITE: PSS0835; IG_LIKE; 1.
 DR PROSITE: PSS0290; IG_MHC; 1.
 KW Immunoglobulin domain; Immunoglobulin C region.
 FT NON TER 1 1
 FT DOMAIN 6 99
 FT DISULFID 27 85
 FT DISULFID 103 103
 SQ SEQUENCE 104 AA; 11254 MW; CE4B67868862D3 CRC64;

Query Match
 Best Local Similarity 4.7%; Score 161; DB 1; Length 104;
 Matches 35; Conservative 15; Mismatches 48; Indels 4; Gaps 2;

QY 521 QPRE-POVYTLPPSRDELTKNOVSLTCLVKGYPSPDIAVESNGOPENYKTPPYLDS 579
 1 QPKSTPLTVFPSPSELKENKATLVCLISNPSGVTAMKANGPITQGVDTSNPTKE 60

QY 580 DGSFELYSKLTVDKSRMOQGVFSCVMEALNHNHTQKSL 621
 61 GNRFMASFLHTSDWRSHNSFTQVTHE--GDTVEKSL 99

Db 61 GNRFMASFLHTSDWRSHNSFTQVTHE--GDTVEKSL 99

RESULT 72
 HB2D_PIG STANDARD; PRT; 258 AA.
 ID HB2D_PIG
 AC P15963;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE SLA class II histocompatibility antigen, DQ haplotype D beta chain precursor.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxID=9623;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90361905; PubMed=2391424;
 RA Gustafsson K., Leguenn C., Hirsch F., Germana S., Pratt K.,
 Sachs D.H.;
 RT "Class II genes of miniature swine. IV. Characterization and
 expression of two allelic class II DQB cDNA clones.";
 RL J. Immunol. 145:1946-1951(1990).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by, and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M31498; AAA31085.1; -
 DR HSSP: P13760; 2SER.
 DR InterPro: IPR007110; Ig-like.
 DR InterPro: IPR003597; Ig_c1.
 DR InterPro: IPR003066; IG_MHC.
 DR InterPro: IPR000353; MHC_II_beta.
 DR Pfam: PF00047; Ig; 1.
 DR Pfam: PF00969; MHC_II_beta; 1.
 DR ProDom: PD000328; MHC_II_beta; 1.
 DR SMART: SM00407; IGc1; 1.
 DR PROSITE: PSS0835; IG_LIKE; 1.
 DR PROSITE: PSS0290; IG_MHC; 1.
 KW MHC II; Transmembrane; Glycoprotein; Signal.
 FT SIGNAL 1 31
 FT CHAIN 32 258
 FT DOMAIN 32 123
 FT DOMAIN 124 217
 FT DOMAIN 218 227
 SLA CLASS II HISTOCOMPATIBILITY ANTIGEN,
 DQ HAPLOTYPE D BETA CHAIN.
 EXTRACELLULAR BETA-1.
 CONNECTING PEPTIDE.

FT TRANSMEM 228 248
 FT DOMAIN 249 258
 FT DISULFID 44 108
 FT DISULFID 146 202
 FT CARBOHYD 48 48
 SQ SEQUENCE 258 AA; 29262 MW; E3AC75110AED47C3 CRC64;

Query Match
 Best Local Similarity 4.7%; Score 161; DB 1; Length 258;
 Matches 46; Conservative 28; Mismatches 80; Indels 8; Gaps 4;

QY 456 WYDVGVEVNAKTKREEQNSTYRVSVTLVHQPMNGKEYCKCVSKALPAPIEKT 515
 54 WSDVR-INYGEPLRFPDSMDGEYRVPFLGRPADYLNQKQKALQKRAELDTVCNHY 112

QY 516 SKAG--QPREPOVYTLPPSRDELTKNOVSLTCLVKGYPSPDIAVESNGOPEN-NYK 571
 113 QIEGTTLQRRVQPTTISPSKAEALNHNHLVCAVTDYFSPQVQVFRNGQETAGVV 172

QY 572 TTPPYLDSGSFELYSKLTVDKSRMOQGVFSCVMEALNHNHTQKSL 613
 173 STPLIRNGD--WTYQVLMELNLRGSDVYTCRVEHSSLQS 211

Db 173 STPLIRNGD--WTYQVLMELNLRGSDVYTCRVEHSSLQS 211

RESULT 73
 VCAM1_HUMAN STANDARD; PRT; 739 AA.
 ID VCAM1_HUMAN
 AC P19320;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Vascular cell adhesion protein 1 precursor (V-CAM 1) (CD106 antigen) (INCM-100).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Umbilical vein;
 RX MEDLINE=91016951; PubMed=1699207;
 RA Polte T., Newman W., Gopal T.V.;
 RT "Pull length vascular cell adhesion molecule 1 (VCAM-1).";
 RL Nucleic Acids Res. 18:5901-5901(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90090619; PubMed=2688898;
 RA Osborn L., Hession C., Tizard R., Vassallo C., Luhowskyj S.,
 RA Chi-Rosco G., Lobb R.;
 RT "Direct expression cloning of vascular cell adhesion molecule 1, a
 RT cytokine-induced endothelial protein that binds to lymphocytes.";
 RL Cell 59:1203-1211(1989).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91352090; PubMed=1715583;
 RA Cybulsky M.I., Fries J.W.U., Williams A.J., Sulten P., Eddy R.,
 RA Byers W., Shows T., Gimbrone M.A. Jr., Collins T.;
 RT "Gene structure, chromosomal location, and basis for alternative mRNA
 RT splicing of the human VCAM1 gene.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:7859-7863(1991).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91201302; PubMed=1707873;
 RA Hession C., Tizard R., Vassallo C., Schiffer S.B., Goff D., Moy P.,
 RA Chi-Rosco G., Luhowskyj S., Lobb R., Osborn L.;
 RT "Cloning of an alternate form of vascular cell adhesion molecule-1
 RT (VCAM1).";
 RL J. Biol. Chem. 266:6682-6685(1991).
 RN [5]
 RP SEQUENCE FROM N.A., AND VARIANTS PHE-318; ALA-384; ALA-413 AND
 RP LEU-716.
 RA Rieder M.J., Armel T.Z., Carrington D.P., Ozuna M., Kuldanek S.A.,

RA Rajkumar N.R., Toth E.J., Yi O., Nickerson D.A.;
 RL Submitted (Aug-2002) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A. (ISOFORM LONG).
 RC TISSUE=Retinal pigment epithelium;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins E.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stepieton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Uedin T.B., Toshiyuki S., Carrinci P., Prange C.,
 RA Rana S.S., Lequellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bonak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunatirne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butcherfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
 RA "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [7]
 RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 25-226.
 RX MEDLINE=95147978; PubMed=7531291;
 RA Jones E.Y., Haxlous K., Bottomley M.J., Robinson R.C., Driscoll P.C.,
 RA Edwards R.M., Clements J.M., Dudgeon T.J., Stuart D.I.;
 RT "Crystal structure of an integrin-binding fragment of vascular cell
 RL adhesion molecule-1 at 1.8-A resolution."
 RN Nature 373:539-544 (1995).
 RN [8]
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 25-220.
 RX MEDLINE=95296382; PubMed=7539925;
 RA Wang J.-H., Pepinsky R.B., Stehle T., Liu J.-H., Karpusas M.,
 RA Browning B., Osborn L.;
 RT "The crystal structure of an N-terminal two-domain fragment of
 RT vascular cell adhesion molecule 1 (VCAM-1): a cyclic peptide based on
 RT the domain 1 C-D loop can inhibit VCAM-1-alpha 4 integrin
 RT interaction."
 RL Proc. Natl. Acad. Sci. U.S.A. 92:5714-5718 (1995).
 RN [9]
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 25-220.
 RA Wang J.-H., Stehle T., Pepinsky R.B., Liu J.-H., Karpusas M.,
 RA Osborn L.;
 RT "Structure of a functional fragment of VCAM-1 refined at 1.9-A
 RT resolution."
 RL Acta Crystallogr. D 52:369-379 (1996).
 CC -1- FUNCTION: IMPORTANT IN CELL-CELL RECOGNITION. APPEARS TO FUNCTION
 CC IN LEUKOCYTE-ENDOTHELIAL CELL ADHESION. INTERACTS WITH THE BETA-1
 CC INTEGRIN VL44 ON LEUKOCYTES, AND MEDIATES BOTH ADHESION AND SIGNAL
 CC TRANSDUCTION. THE VCAM1/VL44 INTERACTION MAY PLAY A
 CC PATHOPHYSIOLOGIC ROLE BOTH IN IMMUNE RESPONSES AND IN LEUKOCYTE
 CC EMIGRATION TO SITES OF INFLAMMATION.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing: Named isoforms=2;
 CC Comment=Additional isoforms seem to exist;
 CC Name=Long;
 CC IsoId=PI9320-1; Sequence=Displayed;
 CC Name=Short;
 CC IsoId=PI9320-2; Sequence=VSP_002580;
 CC -1- TISSUE SPECIFICITY: Expressed on inflamed vascular endothelium, as
 CC well as on macrophage-like and dendritic cell types in both normal
 CC and inflamed tissue.
 CC -1- INDUCTION: By cytokines (e.g. IL-1, TNF-alpha).
 CC -1- PTM: Sialoglycoprotein.
 CC -1- DISEASE: May play an important role in the genesis of
 CC atherosclerosis and rheumatoid arthritis.
 CC -1- SIMILARITY: Contains 7 immunoglobulin-like C2-type domains.

CC -1- DATABASE: NAME=PROV; NOTE=CD guide CD106 entry;
 CC WWW=<http://www.ncbi.nlm.nih.gov/prov/cd/cd106.htm>.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ch/announce/>
 CC or send an email to license@ebi.ebi.ch).
 CC -----
 CC EMBL; X53051; CAA37218.1; -;
 CC EMBL; M30257; AAS1917.1; ALT_TERM.
 CC EMBL; M73255; AAB61270.1; -;
 CC EMBL; M60335; AAA61269.1; -;
 CC EMBL; AF536818; AAM96190.1; -;
 CC EMBL; BC017276; AAH17276.1; -;
 CC PIR; A41288; A41288.
 CC PIR; B41288; B41288.
 CC PDB; 1VCA; 15-SRP-95.
 CC PDB; 1VSC; 20-TUN-96.
 CC PDB; 1IJ9; 07-NOV-01.
 CC Genew; HGNC:12663; VCAM1.
 CC MIM; 192225; -;
 CC InterPro; IPR003987; ICAM VCAM-1.
 CC InterPro; IPR007110; Ig-Like.
 CC InterPro; IPR003598; Ig-C2.
 CC InterPro; IPR003989; VCAM-1.
 CC Pfam; PF00047; Ig; 6.
 CC PRINTS; PR01472; ICAMVCAM1.
 CC PRINTS; PR01474; VCAM1.
 CC SMART; SM00408; IGC2; 3.
 CC PROSITE; PS50835; IG_LIKE; 5.
 CC Immunoglobulin domain; Glycoprotein; Cell adhesion; Transmembrane;
 CC Repeat; Signal; Alternative splicing; Polymorphism; 3D-structure.
 CC SIGNAL 1 24
 CC CHAIN 25 739
 CC TRANS 25 698
 CC DOMAIN 699 720
 CC TRANS 721 739
 CC DOMAIN 721 739
 CC DOMAIN 25 105
 CC DOMAIN 109 212
 CC DOMAIN 223 309
 CC DOMAIN 312 399
 CC DOMAIN 408 506
 CC DOMAIN 511 595
 CC DOMAIN 600 684
 CC DISULFID 47 95
 CC DISULFID 52 99
 CC DISULFID 137 195
 CC CARBOHYD 273 273
 CC CARBOHYD 365 365
 CC CARBOHYD 417 417
 CC CARBOHYD 463 463
 CC CARBOHYD 531 531
 CC CARBOHYD 561 561
 CC CARBOHYD 601 601
 CC VARSPLIC 310 402
 CC VARIANT 318 318
 CC VARIANT 384 384
 CC VARIANT 413 413
 CC VARIANT 716 716
 CC STRAND 26 30
 CC STRAND 34 38
 CC STRAND 39 40
 CC STRAND 43 50
 CC STRAND 56 61
 CC TURN 62 63
 CC STRAND 74 74
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC Missing (in isoform Short).
 CC S -> F.
 CC /FTID=VAR_013309.
 CC T -> A.
 CC /FTID=VAR_014310.
 CC G -> A.
 CC /FTID=VAR_014311.
 CC I -> L.
 CC /FTID=VAR_014312.

FT TURM 75 76
 FT STRAND 77 82
 FT HELIX 87 88
 FT STRAND 92 99
 FT TURM 100 101
 FT STRAND 102 114
 FT STRAND 120 123

Query Match 4.78; Score 160.5; DB 1; Length 739;
 Best Local Similarity 21.58; Pred. No. 0.0081;
 Matches 118; Conservative 91; Mismatches 217; Indels 123; Gaps 27;

8 RHLLVQLALPLPAAT--QGNKVVLGGKDPVELTCTAGSKKSIOPHKMSNGIKILGN 65
 211 KQAVKELQVYISPKNTYISVPSTKIQEGGSVTMTCSSEGLPAPLFW----- 258
 66 GSFLLTKGPKSLNDRASRLMDQGNFLLIKLIKIEDSPYICE---VEDQKEVOLT 121
 259 -----SKKLDNGNLQHL---SGNATLTLIMRMDSGIYCEGVNLIGKRKEVELI 307
 122 V----FGLTANSDTHLQ--GQSILTLLESPGSSPSVQCRSPRGKNIQ-----GKT 168
 308 VQKPFVEISPGRIAAQIGDSVMTLCSVMGCESPSPFSWRTQIDSPISGKVRSEGTNST 367
 169 LSVSOLQLODGTMTCTVLQNOKKVEFKIDIVVLAPOKASSIYK---KEGEVRS--- 222
 368 LTLSPVSENEHSTLCTVTCGHKKLEKGIQVELYSPRDEPIEMSGGLVNGSSVTWSCV 427
 223 ---FPL-AFTVEKLTSGLLMOAERASSKSWITF---DLNKEVSVKRVTDPELQ- 273
 428 PSVYPRDLREIELKGETIL-----ENIEFIEDTMSLEKSKLEMTFPIETED 476
 274 MGKL-----PLHLTPALQVYAGSGLTLALBAKIGKLEHVNLYVMATQLOK---N 325
 477 TGRALVQOAKLHIDMEFEPRKORS---TOTLVYVNAF--RDTVLVSPSIIIEGSSVN 531
 326 LTCEVWGPTSPKMLSLKENKEKAVSKRPVWVNL---PEAGMMOCL-LSDSGOVLLE 381
 532 MTCISQCFPAKTIWKRQPLNGELQPLSENATLTLITKMEGSGVYCEGINQAGSRKE 551
 382 SNIKVLPTWSTPVPCCAPBPKSCDKTHTCPELLGGPSVFLFPP---KPKDTLMIISRT-- 435
 592 VEILIQVT-----PKDKILT-----AFPSVSEVGGTVIISCTCGN 627
 436 -PEVTCVVDVSHEDPEVKRMVYDVY-EVHNAKTRP---EQNSTYRVVSLTVLH 469
 628 VPR-TWIIKKKAKETGDTVLK-SIDGAYTIRKQLKDAGVYECESKVKVQSOLRSLTDV 685
 490 QDMUNGKEY 498
 686 QGRENNKQY 694

RESULT 74
 LAC_HUMAN STANDARD; PRT; 105 AA.
 AC P01842; P80423;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE IG lambda chain C regions.
 GN IGLC1 AND IGLC2 AND IGLC3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE (BENCE-JONES PROTEIN SH).
 RX MEDLINE=70166723; PubMed=4909564;
 RA Titani K., Wikler M., Shinoda T., Putnam F.W.;
 RT "The amino acid sequence of a lambda type Bence-Jones protein. 3. The
 RT complete amino acid sequence and the location of the disulfide
 RT bridges.";

RL J. Biol. Chem. 245:2171-2176 (1970).
 [2]
 RP SEQUENCE (BENCE-JONES PROTEIN X), AND DISULFIDE BONDS.
 RX MEDLINE=69088380; PubMed=4883841;
 RA Milstein C., Clegg J.B., Jarvis J.M.;
 RT "Immunoglobulin lambda-chains. The complete amino acid sequence of a
 RT Bence-Jones protein.";
 RL Biochem. J. 110:631-652 (1966).
 [3]
 RP SEQUENCE (BENCE-JONES PROTEIN NIG-64).
 RX MEDLINE=83186114; PubMed=6404900;
 RA Kametani F., Takayasu T., Suzuki S., Shinoda T., Okuyama T.,
 RA Shimizu A.;
 RT "Comparative studies on the structure of the light chains of human
 RT immunoglobulins. IV. Assignment of a subgroup.";
 RL J. Biochem. 93:421-429 (1983).
 [4]
 RP SEQUENCE (BENCE-JONES PROTEIN KERN).
 RX MEDLINE=71150336; PubMed=5549568;
 RA Ponetning H., Hesse M., Hilschmann N.;
 RT "Structural rule of antibodies. Primary structure of a monoclonal
 RT immunoglobulin-L-chain of the lambda type, subgroup IV (Bence-Jones-
 RT protein Kern). V. The complete amino acid sequence and its genetic
 RT interpretation.";
 RL Hoppe-Seyler's Z. Physiol. Chem. 352:247-266 (1971).
 [5]
 RP PARTIAL SEQUENCE (MYELOMA PROTEIN NEMO).
 RX MEDLINE=74109253; PubMed=4814727;
 RA Chen B.L., Poljak R.J.;
 RT "Amino acid sequence of the (lambda) light chain of a human myeloma
 RT immunoglobulin (IgG New).";
 RL Biochemistry 13:1295-1302 (1974).
 [6]
 RP SEQUENCE (DOT).
 RX MEDLINE=95255298; PubMed=7737190;
 RA Scoppioni M., Bellotti V., Negri A., Merlini G., Garver F., Ferri G.;
 RT "Characterization of the two unique human anti-flavin monoclonal
 RT immunoglobulins.";
 RL Eur. J. Biochem. 228:886-893 (1995).
 [7]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF NEMO.
 RX MEDLINE=75046825; PubMed=4215080;
 RA Poljak R.J., Amzel L.M., Avey H.P., Chen B.L., Phizackerley R.P.,
 RA Saul F.;
 RT "The three-dimensional structure of the fab' fragment of a human
 RT myeloma immunoglobulin at 2.0-A resolution.";
 RL Proc. Natl. Acad. Sci. U.S.A. 71:3440-3444 (1974).
 [8]
 RP PARTIAL SEQUENCE (BENCE-JONES PROTEIN MCG).
 RX MEDLINE=75013804; PubMed=4415202;
 RA Fect J.W., Deutsch H.F.;
 RT "Primary structure of the MCG lambda chain.";
 RL Biochemistry 13:4102-4114 (1974).
 [9]
 RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF MCG.
 RX Edmundson A.B., Ely K.R., Abola E.E., Schiffer M.,
 RA Panagiotopoulos N.;
 RT "Rotational allomerism and divergent evolution of domains in
 RT immunoglobulin light chains.";
 RL Biochemistry 14:3953-3961 (1975).
 [10]
 RP X-RAY CRYSTALLOGRAPHY OF MCG.
 RX MEDLINE=90133913; PubMed=2515285;
 RA Ely K.R., Herron J.N., Harker M., Edmundson A.B.;
 RT "Three-dimensional structure of a light chain dimer crystallized in
 RT water. Conformational flexibility of a molecule in two crystal
 RT forms.";
 RL J. Mol. Biol. 210:601-615 (1989).
 [11]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=82080680; PubMed=6273747;
 RA Hietar P.A., Hollis G.F., Kormeyer S.J., Waldmann T.A., Leder P.;
 RT "Clustered arrangement of immunoglobulin lambda constant region genes

```

RT in man."
RL Nature 294:536-540 (1981).
CC -1- MISCELLANEOUS: The sequence shown is the Kern-/Oz-/MCG- chain
CC found in proteins SH, X, and NIG-64. The Kern protein has the
CC Kern+ marker, the NEMW protein has the Oz+ marker, the MCG protein
CC has the Kern+ marker, and the MCG+ marker.
CC -1- MISCELLANEOUS: Six tandem lambda-type genes were identified and
CC the 5' were sequenced. These correspond to the MCG sequence
CC (lambda-1), the Kern-/Oz- sequence (lambda-2) and the Kern-/Oz+
CC sequence (lambda-3).
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1sb-sib.ch/announce/
CC or send an email to license@1sb-sib.ch).
CC -----
DR EMBL; J00253; AAA59107.1; -.
DR EMBL; L38562; AAB36581.1; ALT_INIT.
DR EMBL; X51754; CAB38569.1; ALT_INIT.
DR EMBL; X51755; CA36049.1; -.
DR PIR; A92057; L2HU.
DR PDB; 2MCG; 15-JUL-92.
DR PDB; 7FAB; 31-JAN-94.
DR PDB; 1AOK; 04-FEB-99.
DR PDB; 1LIL; 15-MAY-97.
DR Genew; HGNC:5855; IGLC1.
DR Genew; HGNC:5856; IGLC2.
DR Genew; HGNC:5857; IGLC3.
DR MIM; 147220; -.
DR GO; GO:0003823; F:antigen binding; NAS.
DR GO; GO:0006955; P:immune response; NAS.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003597; IG-cl.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00407; IGL; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
DR PROSITE; PS00290; IG_MHC; 1.
KW Immunoglobulin domain; Immunoglobulin C region; Bence-Jones protein;
KW 3D-structure.
FT DOMAIN 1 100 IG-LIKE.
FT DISULFID 27 86 INTERCHAIN (WITH HEAVY CHAIN).
FT DISULFID 104 104 A->N (IN MCG+ MARKER).
FT VARIANT 5 5 FTID=VAR_003698.
FT VARIANT 7 7 S->T (IN MCG+ MARKER).
FT VARIANT 45 45 S->G (IN KERN+ MARKER).
FT VARIANT 56 56 T->K (IN MCG+ MARKER).
FT VARIANT 82 82 R->K (IN OZ+ MARKER).
FT VARIANT /FTID=VAR_003901.
FT VARIANT /FTID=VAR_003902.
FT STRAND 8 11
FT HELIX 15 19
FT TURN 20 21
FT STRAND 24 32
FT STRAND 38 43
FT TURN 44 45
FT STRAND 46 48
FT STRAND 52 59
FT TURN 61 62
FT STRAND 65 72
FT HELIX 75 80
FT STRAND 84 89
FT STRAND 94 99
SO SEQUENCE 105 AA; 11236 MM; DCD9C7C201C13CC2 CRC64;

```

```

Query Match 4.7%; Score 159; DB 1; Length 105;
Best Local Similarity 34.1%; Pred. No. 0.00079;
Matches 31; Conservative 18; Mismatches 40; Indels 2; Gaps 2;

Qy 521 QPR-EPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQP-ENNYKTTTPVL 578
Db 1 QPKAPVTLTFPSESELQANKATLVCLISDFYGAVTVMKADSSPVKAGVETTTPSKQ 60

Qy 579 SDGSFPLYSKLTVDKSRWQGNVFCGVME 609
Db 61 SNNRYAASSTYALSLTPBQWKSHRSYSCQVTHE 91

RESULT 75
LAC_PIG STANDARD; PRT; 105 AA.
ID LAC_PIG
AC P01846;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig lambda chain C region.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE.
RX MEDLINE=78000254; PubMed=409425;
RA Novotny J., Franek F., Margolies M.N., Haber E.;
RT "Amino acid sequence of normal (microheterogeneous) porcine
RT Immunoglobulin lambda chains.";
RL Biochemistry 16:3765-3772(1977).
CC -1- MISCELLANEOUS: This chain was obtained from a mixture of normal
CC Immunoglobulins.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
DR PIR; A02129; LIPG.
DR HSSP; P01842; 2MCG.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003597; IG-cl.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00407; IGL; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
DR PROSITE; PS00290; IG_MHC; 1.
KW Immunoglobulin domain; Immunoglobulin C region.
FT DOMAIN 1 100 IG-LIKE.
FT DISULFID 27 86 INTERCHAIN (WITH HEAVY CHAIN).
FT DISULFID 104 104
SO SEQUENCE 105 AA; 11003 MM; 3817AEBD74C396 CRC64;

Query Match 4.6%; Score 156; DB 1; Length 105;
Best Local Similarity 35.2%; Pred. No. 0.0012;
Matches 32; Conservative 15; Mismatches 42; Indels 2; Gaps 2;

Qy 521 QPR-EPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQP-ENNYKTTTPVL 578
Db 1 QPKAPVTLTFPSESELQANKATLVCLISDFYGAVTVMKADSSPVKAGVETTTPSKQ 60

Qy 579 SDGSFPLYSKLTVDKSRWQGNVFCGVME 609
Db 61 SNNRYAASSTYALSLTPBQWKSHRSYSCQVTHE 91

RESULT 76
LACS_MUSSP STANDARD; PRT; 105 AA.
ID LACS_MUSSP
AC P20765;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig lambda-5 chain C region.

```

```
OS Mus spretus (Western wild mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10096;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88196070; PubMed=3129289;
RA Mami F., Cazeneuve P.A., Kindt T.J.;
RL "Conservation of the immunoglobulin C lambda 5 gene in the Mus gene.";
EMBO J. 7:117-122(1988).
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M35582; AAA39152.1; -.
CC HSSP; P01842; 2MCG.
CC InterPro; IPR007110; IG_1like.
CC InterPro; IPR003597; IG_c1.
CC InterPro; IPR003006; IG_MHC.
CC Pfam; PF00047; Ig_1.
CC SMART; SM00407; IgC1; 1.
CC PROSITE; PS50835; IG_LIKE; 1.
CC PROSITE; PS00290; IG_MHC; 1.
CC Immunoglobulin domain; Immunoglobulin C region.
FT NON_TER 1 1
FT DOMAIN 6 100 IG-LIKE.
FT DISULFID 27 86
FT DISULFID 104 104 INTERCHAIN (WITH HEAVY CHAIN).
SQ SEQUENCE 105 AA; 11674 MW; AAB417DF68471A17 CRC64;

Query Match 4.5%; Score 153.5; DB 1; Length 105;
Best Local Similarity 34.0%; Pred. No. 0.0018;
Matches 35; Conservative 20; Mismatches 43; Indels 5; Gaps 3;

QY 521 QPR-EPQVYTLPPSRDELTKNOVSLTCLVKGFPYPSDIIVEMESNGQP-ENNYKTPPVLDGSPF 578
DB 1 QPRSDPLVTLFLPSIKNLQGNKNTLVCLVSEFPGTLVDMKVDGVPVTVGGVETTPQSPKQ 60
QY 579 SDGSFFLYSKLVYDKSRWQGNVFCSCVMEALHNHYTKSL 621
DB 61 TNNKYMWSYLTLLISDQMPHSRYSCTVTHE---GNTVERKSVS 100

RESULT 77
LAC RABBIT STANDARD; PRT; 105 AA.
ID LAC RABBIT
AC P01847;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig lambda chain C region.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE.
RX MEDLINE=82091105; PubMed=6797414;
RA Garcia I., Jaton J.-C.;
RT "The primary structure of the constant region of Basilea-rabbit
RT immunoglobulin lambda-chains.";
RL Biochem. J. 197;177-183(1981).
CC -1- MISCELLANEOUS: This lambda chain expresses the c7 allotypic
CC specificity.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC PIR; A02130; L7RB.
CC HSSP; P01842; 7FAB.
```

```
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003597; IG_c1.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; Ig_1.
DR SMART; SM00407; IgC1; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
DR PROSITE; PS00290; IG_MHC; FALSE NEG.
KW Immunoglobulin domain; Immunoglobulin C region.
FT NON_TER 1 1
FT DOMAIN 6 100 IG-LIKE.
FT DISULFID 27 86
FT DISULFID 104 104 INTERCHAIN (WITH HEAVY CHAIN).
SQ SEQUENCE 105 AA; 11484 MW; B427513272E8663D CRC64;

Query Match 4.5%; Score 153; DB 1; Length 105;
Best Local Similarity 32.7%; Pred. No. 0.0019;
Matches 32; Conservative 17; Mismatches 45; Indels 4; Gaps 2;

QY 525 PQVYTLPPSRDELTKNOVSLTCLVKGFPYPSDIIVEMESNGQP-ENNYKTPPVLDGSPF 583
DB 6 PSVILFPSSSEELKKNKATLVCLIDFYPRYKVMKMGKADGNSVTVGVDTPQSPKSNKK 65
QY 584 FLYSKLVYDKSRWQGNVFCSCVMEALHNHYTKSL 621
DB 66 AASSFLHLTANQWKSQSVTCQVTHE---GHTVERKSLA 100

RESULT 78
LAC RABBIT STANDARD; PRT; 106 AA.
ID LAC RABBIT
AC P01839;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ig kappa-B4 chain C region.
GN K-BAS.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RA Heidmann O., Rougeon F.;
RT "Multiplicity of constant kappa light chain genes in the rabbit
RT genome: a b4b4 homozygous rabbit contains a kappa-bas gene.";
RL EMO J. 2:437-441(1983).
CC -1- MISCELLANEOUS: In Basilea rabbits, the major type of light chain
CC is lambda. The kappa chain shown is a minor component. All other
CC rabbit B allotypes have Cys-64.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; V01241; CAA24558.1; -.
CC EMBL; V00885; -; NOT_ANNOTATED_CDS.
CC PIR; A02121; K4RBS.
CC HSSP; P01842; 7FAB.
CC InterPro; IPR007110; IG_1like.
CC InterPro; IPR003597; IG_c1.
CC InterPro; IPR003006; IG_MHC.
CC Pfam; PF00047; Ig_1.
CC SMART; SM00407; IgC1; 1.
CC PROSITE; PS50835; IG_LIKE; 1.
CC PROSITE; PS00290; IG_MHC; FALSE NEG.
KW Immunoglobulin domain; Immunoglobulin C region.
FT NON_TER 1 1
```

FT DOMAIN 6 99 IG-LIKE.
 FT DISULFID 27 87
 FT DISULFID 106 106 INTERCHAIN (WITH A HEAVY CHAIN).
 SO SEQUENCE 106 AA; 11279 MW; AF9B928DDA853849 CRC64;
 Query Match 4.5%; Score 153; DB 1; Length 106;
 Best Local Similarity 33.7%; Pred. No. 0.0019;
 Matches 31; Conservative 19; Mismatches 36; Indels 6; Gaps 2;
 QY 520 GQPEPQYTLTPRDELTKNOVSLTCLVKGFPSDIAVESNGQPE---NNYKTPP 575
 DB 1 GDPAPSVLLFPSPKELTGATATVCYANKFSPSDITVWKVDGTTQSGSIENSKT--P 58
 QY 576 VLDSGSFPLYSKLTVDKSRQGNVFSQSYV 607
 DB 59 QSPEDNTYSLSTSLTSQAQXNSHSVYTCFV 90
 RESULT 79
 ID CD22 HUMAN STANDARD; PRT: 847 AA.
 AC P20273; O95699; O95701; O95702; Q01665; Q92872; Q92873;
 AC Q9UQA1; Q9UQA8; Q9UQA9; Q9UQA0; Q9Y2A6;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE B-cell receptor CD22 precursor (Ieu-14) (B-lymphocyte cell adhesion molecule) (Bt-CAM) (Siglec-2).
 GN CD22.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RC SEQUENCE FROM N.A. (ISOFORM CD22-BETA), AND VARIANT HIS-639.
 RC TISSUE=Tonsil;
 RC MEDLINE=91086838; PubMed=1985119;
 RA Wilson G.L., Fox C.H., Fauci A.S., Kehrl J.H.;
 RT "CDNA cloning of the B cell membrane protein CD22: a mediator of B-B cell interactions.";
 RT J. Exp. Med. 173:137-146(1991).
 RN [2]
 RC SEQUENCE FROM N.A. (ISOFORMS CD22-ALPHA AND CD22-BETA).
 RC MEDLINE=93267103; PubMed=8496602;
 RA Wilson G.L., Najfeld V., Kozlow E., Menniger J., Ward D.,
 RT Kehrl J.H.;
 RT "Genomic structure and chromosomal mapping of the human CD22 gene.";
 RT J. Immunol. 150:5013-5024(1993).
 RN [3]
 RC SEQUENCE FROM N.A. (ISOFORMS CD22-ALPHA AND CD22-BETA).
 RA Lemerlin J.E., McCreedy P., Adamson A.W., Burkhart-Schultz K.,
 RA Garcia E., Kyle A., Ramirez M., Stillwagen S., Gaines J.,
 RA Dangaran L., Bruce R., Quan G., Montgomery M., Ow D.,
 RA Kobayashi A., Olsen A.O., Carraro A.V.;
 RA Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RC SEQUENCE FROM N.A. (ISOFORM CD22-BETA).
 RC MEDLINE=90231465; PubMed=1691828;
 RA Stamenkovic I., Seed B.;
 RT "The B-cell antigen CD22 mediates monocyte and erythrocyte adhesion.";
 RT Nature 345:74-77(1990).
 RN [5]
 RC SEQUENCE OF 13-137; 139-239; 241-328 AND 418-502 FROM N.A., AND
 RP VARIANTS THR-34; GLU-152; LYS-203; GLY-664; CYS-669 AND ASP-745.
 RX MEDLINE=99180618; PubMed=10079291;
 RA Hata Y., Tsuchiya N., Matsushita M., Shiota M., Hagiwara K.,
 RA Tokunaga K.;
 RT "Identification of the gene variations in human CD22.";
 RT Immunogenetics 49:280-286(1999).
 RL [6]
 RC IMMUNOGENETICS 49:280-286(1999).
 RP SIALIC ACID BINDING.
 RX MEDLINE=93216636; PubMed=8463235;
 RA Powell L.D., Sgroi D., Sjöberg E.R., Stamenkovic I., Varzi A.;

RT "Natural ligands of the B cell adhesion molecule CD22 beta carry
 RT N-linked oligosaccharides with alpha-2,6-linked sialic acids that are
 RT required for recognition.";
 RT J. Biol. Chem. 268:7019-7027(1993).
 RN [7]
 RC INTERACTION WITH PTPN6.
 RX MEDLINE=95343349; PubMed=7618087;
 RA Doody G.M., Jussemont L.B., Delibrias C.C., Matthews R.J., Lin J.,
 RA Thomas M.L., Fearon D.T.;
 RT "A role in B cell activation for CD22 and the protein tyrosine
 RT phosphatase SHP.";
 RT Science 269:242-244(1995).
 RN [8]
 RC INTERACTION WITH LYN; SYK AND PIK3R1/PIK3R2.
 RX MEDLINE=96257803; PubMed=8647200;
 RA Tuscano J.W., Engel P., Tedder T.F., Agarwal A., Kehrl J.H.;
 RT "Involvement of p72syk kinase, p53/56lyn kinase and phosphatidylyl
 RT inositol-3 kinase in signal transduction via the human B lymphocyte
 RT antigen CD22.";
 RT Eur. J. Immunol. 26:1246-1252(1996).
 RN [9]
 RC INTERACTION WITH PTPN6, SYK AND PLCG1.
 RX MEDLINE=96195207; PubMed=8627166;
 RA Law C.L., Sidorenko S.P., Chandran K.A., Zhao Z., Shen S.H.,
 RA Fischer E.H., Clark E.A.;
 RT "CD22 associates with protein tyrosine phosphatase IC, Syk, and
 RT phospholipase C-gamma(1) upon B cell activation.";
 RT J. Exp. Med. 183:547-560(1996).
 RN [10]
 RC REVIEW.
 RX MEDLINE=97288746; PubMed=9143697;
 RA Tedder T.F., Tuscano J., Sato S., Kehrl J.H.;
 RT "CD22, a B lymphocyte-specific adhesion molecule that regulates
 RT antigen receptor signaling.";
 RT Annu. Rev. Immunol. 15:481-504(1997).
 RN [11]
 RC FUNCTION: Mediates B-cell B-cell interactions. May be involved in
 CC the localization of B-cells in lymphoid tissues. Binds sialylated
 CC glycoproteins; one of which is CD45. Preferentially binds to
 CC alpha2,6-linked sialic acid. The sialic acid recognition site can
 CC be masked by cis interactions with sialic acids on the same cell
 CC surface. Upon ligand induced tyrosine phosphorylation in the
 CC immune response seems to be involved in regulation of B cell
 CC antigen receptor signaling. Plays a role in positive regulation
 CC through interaction with Src family tyrosine kinases and may also
 CC act as an inhibitory receptor by recruiting cytoplasmic
 CC phosphatases via their SH2 domains that block signal transduction
 CC through dephosphorylation of signaling molecules.
 CC SUBUNIT: Predominantly monomer of isoform CD22-beta. Also found as
 CC heterodimer of isoform CD22-beta and a shorter isoform. Interacts
 CC with PTPN6/SHP-1, LYN, SYK, PIK3R1/PIK3R2 and PLCG1 upon
 CC phosphorylation. Interacts with GRR2, INP5D and SHC1 upon
 CC INP5D/SHIP, GRB2 and SHC1.
 CC SUBCELLULAR LOCATION: Type I membrane protein.
 CC ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Comment=Additional isoforms seem to exist;
 CC Name=CD22-beta;
 CC IsoId=P20273-1; Sequence=Displayed;
 CC Name=CD22-alpha;
 CC IsoId=P20273-2; Sequence=VSP_002531;
 CC Name=CD22-alpha;
 CC IsoId=P20273-2; Sequence=VSP_002531;
 CC -1- TISSUE SPECIFICITY: B lymphocytes.
 CC -1- DOMAIN: Contains 4 copies of a cytoplasmic motif that is referred
 CC to as the immunoreceptor tyrosine-based inhibitor motif (ITIM).
 CC This motif is involved in downmodulation of cellular responses.
 CC The phosphorylated ITIM motif binds to the SH2 domain of
 CC PTPN6/SHP-1.
 CC -1- PTM: Phosphorylated both on threonine/serine and tyrosine
 CC residues.
 CC -1- PTM: Phosphorylated on tyrosine residues by LYN (By similarity).
 CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily. SIGLEC
 CC (sialic acid binding Ig-like lectin) family.
 CC -1- SIMILARITY: Contains 6 immunoglobulin-like C2-type domains.

FT	CARBOHYD	67	67	N-LINKED (GLCNAC..)	(POTENTIAL).
	Query Match	4.5%;	Score 152.5;	DB 1;	Length 847;
	Best Local Similarity	19.1%;	Pred. No. 0.031;		
	Matches 130;	Conservative 105;	Mismatches 229;	Indels 217;	Gaps 36;
Oy	LIIKNIKEIDSDPYICE-----	VEDOKEEVOLVLTGLTANSPDTLH	133		
Dd	11 LVLEYLEAFSSSKMKVEPHETLYAMEGACWVICIYTRALDDGLEFILFNNPEYNNTSKS	70			
Oy	134 LQGSLLTLLE--SPGSSPSVOCRBPRGNIOGGKITLSVSQLELODSGTWTCTVLONOK	191			
Dd	71 FDGRFLRYESTKDGKVSEQRVOFLGDKNKC-----TLSTIHPLHNDSSG---LGIRMES	123			
Oy	192 KVEEKIDIVL-----AFQKASSIYYKKGEQYE-----FSRPLAFYTEKLT	233			
Dd	124 KTEMMERILTNVSRPPPHILOPREI---QSQSVITLTCLANFSGYGPFI-----	172			
Oy	234 GSGELMWQAERASSKSWITFDLNKNEVSVKRVTOPDKIQMGKKLPDLHTLPQALPOYAG	293			
Dd	173 ---QLQWLIGCVPMROAAVT-----STSLTIKSYFRNSELKFS-----EQMSH	212			
Oy	294 SGNL-TLALEAKTGKL-HOEVLVVMRAOLO-----KULTGEVWGPTS	335			
Dd	213 HGKIVITQLODADAGKFLSNJDVOLNVKHPTKEIKYTPSDAIVREGDSVTMTCEV-SSSN	271			
Oy	336 PKLM-----LSLKLENKAAYSKREKRVWVLNPAGMCQCLSD-----SGQVLLR	381			
Dd	272 PEYTVWSMLDQISLKKQN-TFLTLNIRE-----VTHQSGKYCCQOVANDVGPGSBEEVFQ	326			
Oy	382 -----SNIKVL-----FWSTSTVPFC-----PAPEPKSCDKTH----	408			
Dd	327 VOYAPESPSTVOIILHPBAVESGOVEFLCMSIANBLPNITYWHNGKMEQGRTBEKVHI	386			
Oy	409 -----TCPEELG-----GPSVFL----FPKPCKDTLMISRTP-----EVCGVVDVSH	447			
Dd	387 LPMHAGTYSVAENILISTGQRGFGAELDVOYPFKATTVIQNMPIRESGTVLISCYNIS	446			
Oy	448 EDEPV-KFNMYVDGVEVHNNAKTTPREBYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNK-	505			
Dd	447 SNPSVTRYEW-----KPHGAMEEPSS-----LGVLIKIQVGMDN-TTICAACNSW	490			
Oy	506 -ALPAPIEKITSKAGQRPREPQYTYLTPRSDELTKNOVSLTCLKVGPSPDIADVEMESG	564			
Dd	491 CSWASPPALNVQYA---PRDVRARKIKIPLEISHGNSVSIQQCPSSHPEXEPVPFEKNG	547			
Oy	565 ----OENNRYKTTTPYLVDGSGFFLYSKLTVDKSRHWOCQGNVSCSYVHEALNHNYTOKS	619			
Dd	548 RLKGESQLMFDSISP-EDAGSY-----SCVVNNSIGQTASKAVTLFVLVAPRR	595			
Oy	620 L--SLSPGLQLDL-----TC	632			
Dd	596 LRVSMSPGDQMEGKSATLTC	616			
RESULT 80					
DTC_HUMAN STANDARD; PRT; 383 AA.					
AC	P0180;				
DT	21-JUL-1986 (Rel. 01, Created)				
DT	21-JUL-1986 (Rel. 01, Last sequence update)				
DT	10-OCT-2003 (Rel. 42, Last annotation update)				
GN	IGH delta chain C region.				
OS	IGHD.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
OX	NCBI_Taxid=9606;				
RN	[1]				
RA	SEQUENCE OF 1-162 (MYELOMA PROTEIN WAH).				
RT	MEDLINE=62082419; PubMed=6947220;				
RT	Putnam F.W., Takahashi N., Tetaert D., Debuire B., Lin L.-C.;				
RT	"Amino acid sequence of the first constant region domain and the				

	RT	hinge region of the delta heavy chain of human IgD."
	RN	Proc. Natl. Acad. Sci. U.S.A. 78:6168-6172(1981).
	RP	[2]
	RA	SEQUENCE OF 158-383 (MYELOMA PROTEIN WAH).
	RX	MEDLINE=81223768; PubMed=6787589;
	RY	Liu L.-C., Putnam F.W.;
	SA	"Primary structure of the Cc region of human immunoglobulin D:
	SB	implications for evolutionary origin and biological function.";
	SC	Proc. Natl. Acad. Sci. U.S.A. 78:504-508(1981).
	SD	[3]
	SE	SEQUENCE OF 103-137 (MYELOMA PROTEIN NIG-65), AND CARBOHYDRATES.
	SF	MEDLINE=82231401; PubMed=7092891;
	SG	Takayasu T., Suzuki S., Kamezani F., Takahashi N., Shinoda T.,
	SH	Okuyama T., Munkata E.; galactosamine-containing glycopeptides in the
	SI	"Amino acid sequence of a human immunoglobulin D,"
	SJ	hinge region of a human immunoglobulin D."
	SK	Biochem. Biophys. Res. Commun. 105:1066-1071(1982).
	SL	[4]
	SM	SEQUENCE OF 158-383 (MYELOMA PROTEIN NIG-65).
	SN	MEDLINE=81199406; PubMed=6786754;
	SO	Shinoda T., Takahashi N., Takayasu T., Okuyama T., Shimizu A.;
	SP	"Complete amino acid sequence of the Cc region of a human delta
	SR	chain."
	SS	Proc. Natl. Acad. Sci. U.S.A. 78:785-789(1981).
	ST	[5]
	SU	CARBOHYDRATE-LINKAGE SITES.
	SV	MEDLINE=81133632; PubMed=708791;
	SW	Takayasu T., Takahashi N., Shinoda T.;
	SX	"Amino acid sequence and location of the three glycopeptides in the
	SY	Fc region of human immunoglobulin D."
	TA	Biochem. Biophys. Res. Commun. 97:635-641(1980).
	TB	-I-STIMILARITY: Contains 3 Immunoglobulin-like domains.
	TC	PIR: A02175; DHHD.
	TD	Genew; HGNC:5480; IGHD.
	TE	MIM: 147170; -
	TF	GO: GO:0003823; F:antigen binding; NAS.
	TG	GO: GO:0006955; P:immune response; NAS.
	TH	InterPro: IPR007110; IG-like.
	TI	InterPro: IPR003597; Ig-cl.
	TJ	InterPro: IPR003006; Ig_MHC.
	TK	Pfam: PF00047; Ig_3.
	TL	SMART: SMO0407; Igcl; 3.
	TM	PROSITE: PS50835; IG_LIKE; 3.
	TO	PROSITE: PS00290; IG_MHC; 2.
	TP	Immunoglobulin domain; Immunoglobulin C region; Glycoprotein; Repeat.
	TQ	NON_TER 1
	TR	DOMAIN 6 98
	TS	DOMAIN 175 263
	TT	DOMAIN 267 373
	TU	INTERCHAIN (WITH LIGHT CHAIN)
	TV	DISULFD 15 15
	TW	DISULFD 28 84
	TX	INTERCHAIN (WITH HEAVY CHAIN)
	TY	DISULFD 161 161
	TZ	DISULFD 109 109
	UA	CARBOHYD 110 110
	UB	CARBOHYD 113 113
	UC	O-LINKED.
	UD	O-LINKED.
	UE	O-LINKED.
	UF	O-LINKED.
	UG	O-LINKED.
	UH	O-LINKED.
	UI	O-LINKED.
	UJ	O-LINKED.
	UK	O-LINKED.
	UL	O-LINKED.
	UM	O-LINKED.
	UN	O-LINKED.
	UU	O-LINKED.
	UV	O-LINKED.
	UW	O-LINKED.
	UX	O-LINKED.
	UY	O-LINKED.
	UZ	O-LINKED.
	VA	O-LINKED.
	VB	O-LINKED.
	VC	O-LINKED.
	VD	O-LINKED.
	VE	O-LINKED.
	VF	O-LINKED.
	VG	O-LINKED.
	VH	O-LINKED.
	VI	O-LINKED.
	VJ	O-LINKED.
	VK	O-LINKED.
	VL	O-LINKED.
	VM	O-LINKED.
	VN	O-LINKED.
	VO	O-LINKED.
	VP	O-LINKED.
	VQ	O-LINKED.
	VR	O-LINKED.
	VS	O-LINKED.
	VT	O-LINKED.
	VV	O-LINKED.
	VG	O-LINKED.
	VH	O-LINKED.
	VI	O-LINKED.
	VJ	O-LINKED.
	VK	O-LINKED.
	VL	O-LINKED.
	VM	O-LINKED.
	VN	O-LINKED.
	VO	O-LINKED.
	VP	O-LINKED.
	VQ	O-LINKED.
	VR	O-LINKED.
	VS	O-LINKED.
	VT	O-LINKED.
	VV	O-LINKED.
	VG	O-LINKED.
	VH	O-LINKED.
	VI	O-LINKED.
	VJ	O-LINKED.
	VK	O-LINKED.
	VL	O-LINKED.
	VM	O-LINKED.
	VN	O-LINKED.
	VO	O-LINKED.
	VP	O-LINKED.
	VQ	O-LINKED.
	VR	O-LINKED.
	VS	O-LINKED.
	VT	O-LINKED.
	VV	O-LINKED.
	VG	O-LINKED.
	VH	O-LINKED.
	VI	O-LINKED.
	VJ	O-LINKED.
	VK	O-LINKED.
	VL	O-LINKED.
	VM	O-LINKED.
	VN	O-LINKED.
	VO	O-LINKED.
	VP	O-LINKED.
	VQ	O-LINKED.
	VR	O-LINKED.
	VS	O-LINKED.
	VT	O-LINKED.
	VV	O-LINKED.
	VG	O-LINKED.
	VH	O-LINKED.
	VI	O-LINKED.
	VJ	O-LINKED.
	VK	O-LINKED.
	VL	O-LINKED.
	VM	O-LINKED.
	VN	O-LINKED.
	VO	O-LINKED.

```

Db      71 STPLQGMKQG-EYKVVGHHTSKSKKEIFRMPESPRAQASVPTNPQPAEGSLAKTTAP 129
Qy      511 I-----EKTISKAKGQPREPO-----VYTLPPSRDEL-TKNQVSLT 545
Db      130 ATRNTATGSGEKKKKKEKEBEDEHETKTPPECPSHTQPLGVYLLPAPVDLRLDKATFT 189
Qy      546 CLVKGFFYSIDIAVWESNGQ-PENNYKTTPTPLVSDGSEFFLYSKLITVDKSRMQGQNVSC 604
Db      190 CFVWSGDKDMLTWEVAKVPTGVEBGLLERSHNGSOSQSHSRLLTLPRLSNAGTSVYC 249
Qy      605 SVMGEAL 611
Db      250 TLNHPSL 256

RESULT 81
KACA_RAT
ID_KACA_RAT STANDARD; PRT; 106 AA.
AC P01836;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig kappa chain C region, A allele.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DA;
RX MEDLINE=82082587; PubMed=6273908;
RA Shepard H.W., Guttman G.A.;
RT "Allelic forms of rat kappa chain genes: evidence for strong
RL selection at the level of nucleotide sequence.";
RT Proc. Natl. Acad. Sci. U.S.A. 78:7064-7068(1981).
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
DR PIR, A02118; K1RTA.
DR HSSP, P01842; 2MCG.
DR InterPro, IPR007110; IG-like.
DR InterPro, IPR003597; IG cl.
DR InterPro, IPR003006; IG_MHC.
DR Pfam, PF00047; Ig_1.
DR SMART, SM00407; IGc1_1.
DR PROSITE, PS50835; IG LIKE; 1.
DR PROSITE, PS00290; IG_MHC; 1.
KW immunoglobulin domain; Immunoglobulin C region.
FT NON TER 1
FT DOMAIN 5 102 IG-LIKE.
FT DISULFID 26 86
FT DISULFID 106 106 INTERCHAIN (WITH A HEAVY CHAIN).
SQ SEQUENCE 106 AA; 11732 MW; B7B12DD9700DD66 CRC64;

Query Match 4.4%; Score 150; DB 1; Length 106;
Best Local Similarity 31.5%; Pred. No. 0.003;
Matches 29; Conservative 23; Mismatches 28; Indels 12; Gaps 2

Qy 525 PQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVWESNGQPENNYKTTPTPLDS----- 579
Db 5 PTVAIFPPSMQQLTSGATVVCFVNPFPRDISVWKIKDGSQRD-----GVLDSTVDQD 59
Qy 580 --DGSFFLYSKLITVDKSRMQGQNVSCVME 609
Db 60 SKDSTYSMSITLSLTKEVERHNLYTCEVHK 91

RESULT 82
DCC_MOUSE
ID_DCC_MOUSE STANDARD; PRT; 1447 AA.
AC P70211;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)

```

DE Tumor suppressor protein DCC precursor.
 GN DCC.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 ON NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS A; B AND C).
 RC STRAIN=BALB/c; TISSUE=Brain;
 RA MEDLINE=96112625; PubMed=8570174;
 RX Cooper H.M., Ames P., Brito J., Gad J., Wilks A.F.;
 RT "Cloning of the mouse homologue of the deleted in colorectal cancer
 gene (mDCC) and its expression in the developing mouse embryo.";
 RL Oncogene 11:2243-2254(1995).
 RN [2]
 RP REVISIONS.
 RC STRAIN=BALB/c; TISSUE=Brain;
 RA Cooper H.M.;
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: IMPLICATED AS A TUMOR SUPPRESSOR GENE.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=A;
 CC IsoId=PF0211-1; Sequence=Displayed;
 CC Note=Isoform B is produced by alternative initiation at Met-85
 CC of Isoform A;
 CC Name=C;
 CC IsoId=PF0211-2; Sequence=VSP_002501;
 CC Event=Alternative initiation;
 CC Comment=2 isoforms, A (shown here) and B, are produced by
 CC alternative initiation at Met-1 and Met-85;
 CC -1- TISSUE SPECIFICITY: In the embryo, expressed at high levels in the
 CC developing brain and neural tube. In adult, highly expressed in
 CC brain with very low levels found in testis, heart and thymus.
 CC Isoform C is expressed only in the embryo.
 CC -1- DEVELOPMENTAL STAGE: Low levels in early gestation. Highest levels
 CC expressed during mid gestation. Levels decrease in late gestation
 CC and remain at this level in the adult.
 CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily. DCC family.
 CC -1- SIMILARITY: Contains 4 immunoglobulin-like C2-type domains.
 CC -1- SIMILARITY: Contains 6 fibronectin type III domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X85788; CAAS9786.1; -.
 DR HSSP; P56276; ITLC.
 DR MGD; MGI:94869; DCC.
 DR InterPro; IPR008957; FN III-like.
 DR InterPro; IPR003961; FN-III.
 DR InterPro; IPR003962; FNIII_subd.
 DR InterPro; IPR007110; IG-like.
 DR InterPro; IPR003598; IG_c2.
 DR Pfam; PF00041; fn3; 6.
 DR Pfam; PF00047; Ig; 4.
 DR PRINTS; PR00014; FNTPETIII.
 DR SMART; SM00060; FN3; 6.
 DR SMART; SM00408; IGC2; 3.
 DR PROSITE; PS50835; IG LIKE; 4.
 KW Glycoprotein; Immunoglobulin domain; Transmembrane; Signal;
 KW Repeat; Anti-oncogene; Alternative initiation; Alternative splicing.
 FT SIGNAL 1 25
 FT CHAIN 26 1447
 FT INIT MET 85 1447
 FT DOMAIN 26 1097
 FT TRANSMEM 1098 1122
 FT POTENTIAL.
 FT TUMOR SUPPRESSOR PROTEIN DCC, ISOFORM A.
 FT FOR ISOFORM B.
 FT EXTRACELLULAR (POTENTIAL).
 FT POTENTIAL.

FT DOMAIN 1123 1447 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 36 135 IG-LIKE C2-TYPE 1.
 FT DOMAIN 139 229 IG-LIKE C2-TYPE 2.
 FT DOMAIN 234 326 IG-LIKE C2-TYPE 3.
 FT DOMAIN 331 416 IG-LIKE C2-TYPE 4.
 FT DOMAIN 426 522 FIBRONECTIN TYPE-III 1.
 FT DOMAIN 525 618 FIBRONECTIN TYPE-III 2.
 FT DOMAIN 619 716 FIBRONECTIN TYPE-III 3.
 FT DOMAIN 722 816 FIBRONECTIN TYPE-III 4.
 FT DOMAIN 840 940 FIBRONECTIN TYPE-III 5.
 FT DOMAIN 941 1042 FIBRONECTIN TYPE-III 6.
 FT DISULFID 61 117 BY SIMILARITY.
 FT DISULFID 161 212 BY SIMILARITY.
 FT DISULFID 261 310 BY SIMILARITY.
 FT DISULFID 352 400 BY SIMILARITY.
 FT CARBOHYD 60 60 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 94 94 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 318 318 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 478 478 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 628 628 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 702 702 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 819 838 Missing (in isoform C).
 FT FTId=VSP_002501.
 SQ SEQUENCE 1447 AA; 158298 MW; 0D1F1097C22D5B9F CRC64;
 Query Match 4.4%; Score 149.5; DB 1; Length 1447;
 Best Local Similarity 19.0%; Pred. No. 0.098; Indels 275; Gaps 32;
 Matches 134; Conservative 79; Mismatches 219;
 QY 34 GDTVELTCTASQKSIQFHKNSNQIKILGNQSFELTKGPKSLND-RADSRRLMDQGNF 92
 DB 154 GDTVLKKEVISEPPTTHMQNQ-----DLNPLPGSRVVLPSG-- 195
 QY 93 PLIINKLKIEDSYICEVED-----QKEVQLL-----VFGLTNSDTHLQOQ 137
 DB 196 ALQISRLQPGDGVRCSPARNPASIRTGNEAEVRLISPGILRQVLFQRPANVIAIEGK 255
 QY 138 SLTL--TLSEPPG-----SPVQCRSPRKNIQGGKTLVSQLELSDGTMCTVQ 188
 DB 256 DAVLECCVSGYPPEFTMLRGEVQLNSKR-VSLQGSNLLISVTDSDSTYTCVVY 314
 QY 189 NQKVEFKIDIVLA--FOKASIVYKKEGEQVFPFLAFTVEKLTSGELMQAERA 245
 DB 315 KRENISASAEVLVLPWFLNHPNLVYAESNDIEF-----EQA 353
 QY 246 SSSKSWTFD-LKNKEVSKRVTDPKLQWKKPLPLTLPLQALPOVAGSNLTL----- 299
 DB 354 VSGKPVPTVMNMGNDV-----VIPADYFQIVGGSNLRILGVYK 392
 QY 300 -----ALEAKTGLHGVNVLVVMRATQQLNLTCEVWGPTSPRLMLSLKLENKAV 351
 DB 393 SDEGFYQVAENEAQNASQALIVPKAIPSSIL--PSAPDVLPLVSSRFVRL 447
 QY 352 SKR-----EKPVVLNPEA-----CM 367
 DB 448 SRRPRAEAKGNITFTVFRREGDNRERALNTTQGSLOLYGNLKEPMYTRVAVANE 507
 QY 368 W-----QLLSDSGV-----LLESNIKVLFTWSTPYPCAPPE--KSCDK 406
 DB 508 WGRGESSQPIKATQDELQVGPVENLHVAVSTSPSILITWEPAPAVANGVGRFLCTE 567
 QY 407 THTCPE-----LLG-----GPSVLPFPKPKDITMISRTPEV 438
 DB 568 VSTGKEQNIENVGLSYKLEGLKFKTEYTLFLAYNRVGGV-----STDIT 613
 QY 439 TCYVV-DVSHDEPE-----VKFNM-----YVDGVEHNAKTPREE--- 473
 DB 614 TVVTLSDVSAAPPQNSLEVANRSIKVSWLPSPSTQNGFLTGKYLIRKRTKREMER 673
 QY 474 -QNSTYRVVSVLTVLHODMLNGEKKCVSNKALPAPIEKTIISKAGQPREQVYTLPL 532
 DB 674 LRENNLMVLTGLE-----KSGQYSFOVSAMTV-----NOTGPP--SNWYTAET 715

OY 533 SDELTKNQVSLTCLVKGYPSDIWESNGCPENN---YKTPPV 576
 DB 716 PENIDDESQVP-----DQPSLHV-----RPQTNCLIMSWTEPL 749
 RESULT 83
 CMLL_HUMAN
 ID CAML_HUMAN STANDARD; PRT; 1257 AA.
 AC P32004; Q8TA87;
 DT 01-JUL-1993 (Rel. 26, Last Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Neural cell adhesion molecule L1 precursor (N-CAM L1) (CD171 antigen).
 GN L1CAM OR CAML1 OR MIC5.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OK NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92031698; PubMed=1932117;
 RA Kobayashi M., Miura M., Asou H., Uyemura K.;
 RT "Molecular cloning of cell adhesion molecule L1 from human nervous
 tissue: a comparison of the primary sequences of L1 molecules of
 RT different origin.";
 RL Biochim. Biophys. Acta 1090:238-240(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=92120663; PubMed=1769655;
 RA Hlavin M.L., Lemmon V.;
 RT "Molecular structure and functional testing of human L1CAM: an
 RT interspecies comparison.";
 RL Genomics 11:416-423(1991).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92329299; PubMed=1627459;
 RA Reid R.A., Hemperly J.J.;
 RT "Variants of human L1 cell adhesion molecule arise through alternate
 RT splicing of RNA.";
 RL J. Mol. Neurosci. 3:127-135(1992).
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Rosenthal A., Coutelle O., Drescher B.;
 RL Submitted (Apr-1994) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97432815; PubMed=9286695;
 RA Brenner V., Nyakatura G., Rosenthal A., Platzter M.;
 RT "Genomic organization of two novel genes on human Xq28: compact head
 RT to head arrangement of IDH gamma and TRAP delta is conserved in rat
 RT and mouse.";
 RL Genomics 44:8-14(1997).
 RN [6]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RX MEDLINE=98147998; PubMed=9479034;
 RA Coutelle O., Nyakatura G., Taudien S., Elgar G., Brenner S.,
 RA Platzter M., Drescher B., Joutel M., Kenwright S., Rosenthal A.;
 RT "The neural cell adhesion molecule L1: genomic organization and
 RT differential splicing is conserved between man and the pufferfish
 RT Fugu.";
 RL Gene 208:7-15(1998).
 RN [7]
 RP SEQUENCE OF 20-36.
 RX MEDLINE=88298876; PubMed=3136168;
 RA Wolff J.M., Frank R., Mujoo K., Spico R.C., Reisfeld R.A.,
 RA Rachej F.G.;
 RT "A human brain glycoprotein related to the mouse cell adhesion
 RT molecule L1.";
 RL J. Biol. Chem. 263:11943-11947(1988).
 RN [8]
 RP SEQUENCE OF 332-371 FROM N.A.

RX MEDLINE=90353957; PubMed=2387585;
 RA Djabali M., Mattei M.-G., Nguyen C., Roux D., Demengeot J.,
 RA Denizot F., Moos M., Schachner M., Goridis C., Jordan B.R.;
 RT "The gene encoding L1, a neural adhesion molecule of the
 RT immunoglobulin family, is located on the X chromosome in mouse and
 RT man.";
 RL Genomics 7:587-593(1990).
 RN [9]
 RP SEQUENCE OF 353-1176 FROM N.A.
 RC TISSUE=Fetal brain;
 RX MEDLINE=92020233; PubMed=1923824;
 RA Rosenthal A., Mackinnon R.N., Jones D.S.C.;
 RT "PCR walking from microdissection clone M54 identifies three exons
 RT from the human gene for the neural cell adhesion molecule L1
 RT (CAM-L1).";
 RL Nucleic Acids Res. 19:5395-5401(1991).
 RN [10]
 RP SEQUENCE OF 809-1257 FROM N.A.
 RC TISSUE=Pancreas;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haile F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ueda T.B., Tomihataki S., Carninci P., Prange C.,
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultky S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko I., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalek U., Smallie D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [11]
 RP SEQUENCE OF 1030-1257 FROM N.A.
 RX MEDLINE=91132183; PubMed=1993895;
 RA Harper J.R., Prince J.T., Healy P.A., Stuart J.K., Nauman S.J.,
 RA Stalcup W.B.;
 RT "Isolation and sequence of partial cDNA clones of human L1: homology
 RT of human and rodent L1 in the cytoplasmic region.";
 RL J. Neurochem. 56:797-804(1991).
 RN [12]
 RP VARIANT HSAS TYR-264.
 RX MEDLINE=94004956; PubMed=8401576;
 RA Joutel M., Rosenthal A., Macfarlane J., Kenwright S., Donnai D.;
 RT "A missense mutation confirms the L1 defect in X-linked hydrocephalus
 RT (HSAS).";
 RL Nat. Genet. 4:331-331(1993).
 RN [13]
 RP VARIANT HSAS/MASA LEU-1194.
 RX MEDLINE=95187172; PubMed=7881431;
 RA Friesen E., Schrander-Stumpel C., Vits L., Coucke P., van Camp G.,
 RA Williams P.J.;
 RT "X-linked hydrocephalus and MASA syndrome present in one family are
 RT due to a single missense mutation in exon 28 of the L1CAM gene.";
 RL Hum. Mol. Genet. 3:2255-2256(1994).
 RN [14]
 RP VARIANTS HSAS GLN-184 AND ARG-452, AND VARIANT MASA GLN-210.
 RX MEDLINE=95004608; PubMed=7920659;
 RA Joutel M., Rosenthal A., Armstrong G., Macfarlane J., Stevenson R.,
 RA Paterson J., Metzger A., Ionescu V., Temple K., Kenwright S.;
 RT "X-linked spastic paraplegia (SPG1), MASA syndrome and X-linked
 RT hydrocephalus result from mutations in the L1 gene.";
 RL Nat. Genet. 7:402-407(1994).
 RN [15]
 RP VARIANTS MASA GLN-210 AND ASN-598.

RX MEDLINE=95004609; PubMed=7920660;
 RA Vits L., van Camp G., Coucke P., Franssen E., de Boule K.,
 RA Reyniers E., Korn B., Poustra A., Wilson G., Schrander-Stumpel C.,
 RA Winter R.M., Schwartz C., Willems P.J.;
 RT "MASA syndrome is due to mutations in the neural cell adhesion gene
 RT LICAM.";
 RL Nat. Genet. 7:408-413 (1994).
 RP [16]
 RP VARIANTS HSAS/MASA SER-9; SER-121; LYS-309; PHE-768; LEU-941 AND
 RP CYS-1070.
 RX MEDLINE=95282776; PubMed=7762552;
 RA Jouet M., Moncla A., Paterson J., McKeown C., Fryer A., Carpenter N.,
 RA Holmberg E., Medelius C., Kenwright S.;
 RT "New domain of neural cell-adhesion molecule L1 implicated in
 RT X-linked hydrocephalus and MASA syndrome.";
 RL Am. J. Hum. Genet. 56:1304-1314 (1995).
 RP [17]
 RP VARIANTS HSAS/MASA GLN-184; GLN-210; TYR-264; ARG-452; ASN-598 AND
 RP LEU-1194.
 RX MEDLINE=96153146; PubMed=8556302;
 RA Franssen E., Lemmon V., van Camp G., Vits L., Coucke P., Willems P.J.;
 RT "CRASH syndrome: clinical spectrum of corpus callosum hypoplasia,
 RT retardation, adducted thumbs, spastic paraparesis and hydrocephalus
 RT due to mutations in one single gene, L1.";
 RL Eur. J. Hum. Genet. 3:273-284 (1995).
 RP [18]
 RP ERRATUM.
 RA Franssen E., Lemmon V., van Camp G., Vits L., Coucke P., Willems P.J.;
 RL Eur. J. Hum. Genet. 4:126-126 (1996).
 RP [19]
 RP VARIANTS HSAS/MASA/SPG1 SER-179 AND ARG-370.
 RX MEDLINE=96057511; PubMed=7562969;
 RA Ritz J.C., Cuppens H., Legius E., Fryns J.-P., Glover T., Marynen P.,
 RA Caessman J.-J.;
 RT "Mutations in L1-CAM in two families with X linked complicated
 RT spastic paraplegia, MASA syndrome, and HSAS.";
 RL J. Med. Genet. 32:549-552 (1995).
 RP [20]
 RP VARIANTS HSAS CYS-194 AND LEU-240.
 RX MEDLINE=97083370; PubMed=8929944;
 RA Gu S.-M., Orth U., Veeke A., Enders H., Klunder K., Schlosser M.,
 RA Engel W., Schwinger E., Gal A.;
 RT "Five novel mutations in the LICAM gene in families with X linked
 RT hydrocephalus.";
 RL J. Med. Genet. 33:103-106 (1996).
 RP [21]
 RP VARIANTS HSAS GLN-184; 439-VAL--THR-443 DEL; CYS-784 AND
 RP 936-LEU--LEU-948 DEL.
 RX MEDLINE=97338664; PubMed=9195224;
 RA Macfarlane J.R., Du J.-S., Peygs M.E., Ramsden S., Domai D.,
 RA Charlton R., Garrett C., Tolmie J., Yates J.R.W., Berry C., Goudie D.,
 RA Moncla A., Lunt P., Hodgson S., Jouet M., Kenwright S.;
 RT "Five novel L1 CAM mutations in families with X-linked
 RT hydrocephalus.";
 RL Hum. Mutat. 9:512-518 (1997).
 RP [22]
 RP VARIANTS HSAS/MASA ASP-691; ARG-698 AND PRO-935.
 RX MEDLINE=98180721; PubMed=9521424;
 RA Du Y.-Z., Srivastava A.K., Schwartz C.E.;
 RT "Multiple exon screening using restriction endonuclease
 RT fingerprinting (REF): detection of six novel mutations in the L1 cell
 RT adhesion molecule (LICAM) gene.";
 RP Query Match 4.3%; Score 148.5; DB 1; Length 1257;
 RP Best Local Similarity 19.5%; Pred. No. 0.094;
 RP Matches 129; Conservative 88; Mismatches 257; Indels 189; Gaps 30;
 QY 8 RHLLV---LQALLPA--ATGCKVVLGGKDTVELTTSQKSGIOGHWKNSQIKI 61
 DB 409 RHGILLANAYIVVQDLPAKILTDNQYMAVVGSTAYLLCKAGCAPVSVQWMLDEGDTTV 468
 QY 62 LGGSGFLTGSPSLNDRASRSRLWQGNFLLIKKLKLEDDDTYICEVEDQKEFVOLL 121

DB 469 IQDERFF-----PYANGTLGIRDLQANDTGRVFCILANDONNVTIM 509
 QY 122 VEGILANDDTHLLO-----GSLTLTLLES--PQSSPSVQGRSPRGKNIQ----- 164
 DB 510 A-NLVKQATQTLQPPRSTTEKXSRVFTTCASDPSPLOPSITWKG-GRLOQLGDSO 567
 QY 165 -----GKTLVSQLELODSGTWCTCTVLQNCXKVEFKIDIVLAF-----QKASSIVYKK 214
 DB 568 KPIEDGR-LVHSHLDYDQGNVSCVASTELDVESRQGLLVGSRGPRVLVLSLHL 626
 QY 215 EGEQVEFSPFLA-----FTVEKLTSGSELWQAEARASSKSWITFDLKNKEVS 263
 DB 627 TOSQVRVSWSPADENHAPLEKTDIEFDKEMAPKMYSLGKVPNGQSTTLTLSPYVHT 686
 QY 264 KVTQDPKIQMKKPLNHLTL--PQALPQ-----YAGSGNLLALBAKTKGLHGVNLYV 316
 DB 687 FVTALINKYGPSPVSETVTPPAAPKPNVDVKBGENET-----TNVYI 733
 QY 317 MRATQLOKRLTCEVGP-----TSPKMLSLKLENKAKVSKREKPVVVLNPEAGMQ- 369
 DB 734 -----TKPLRMMDMNAPOVQYRQ-----WRPGTRGPQGE 765
 QY 370 CLLSDSGOVLESNIKVLPTWSTPVC-----PADEPSC-----DKTHTCPBLIG-- 415
 DB 766 QIVSD--PFLVVSNTSTFVPEIKVQAVNSGKGPEPQVTIGSGEDYPQALPELEGIEI 823
 QY 416 -GPSVFLPPPKDPLMIS--RTPEYTCVVVDVSHEDPEVFNNYVGVGVHNAKTKR 471
 DB 824 LNSSAVLVYKMRPVDLAQVYGLRGYNVT-----TWRESQKHSKRIH 867
 QY 472 EEOY-----NSTRVVSVLTVLHQMVLNGEKYCKVSNKALPAPIEKTISKAGPREPOV 527
 DB 868 KHVYVPANTTISVILSGLRPYSYHLEVOAFNGRSGRA-----SETFTPEGVCHPRA 923
 QY 528 YTL-----PGRDELTKNOVSLTGLVXGFPYSDIAVWESNGQENNYKTP 574
 DB 924 LHLGCSNTSLLRMQP--LSHNGV-LTGIVLSYHPLD-----EGGKQLSPFLRD 972
 QY 575 PVL 577
 DB 973 PEL 975
 RESULT 84
 KAC_HUMAN
 ID KAC_HUMAN STANDARD; PRT; 106 AA.
 AC P01834;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE 19 kappa chain C region.
 GN IGKC.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE (MYELOMA PROTEIN EU).
 RX MEDLINE=71064023; PubMed=5489770;
 RA Gottlieb P.D., Cunningham B.A., Rutishauser U., Edelman G.M.;
 RT "The covalent structure of a human gamma G-immunoglobulin. VI. Amino
 RT acid sequence of the light chain.";
 RL Biochemistry 9:3155-3161 (1970).
 RN [2]
 RP DISULFIDE BONDS.
 RX MEDLINE=71064027; PubMed=4923144;
 RA Gall W.E., Edelman G.M.;
 RT "The covalent structure of a human gamma G-immunoglobulin. X.
 RT Intrachain disulfide bonds.";
 RL Biochemistry 9:3188-3196 (1970).
 RN [3]
 RP SEQUENCE (BENCE-JONES PROTEIN TI).
 RX MEDLINE=72188439; PubMed=5027703;

RA Suter L., Barnikol H.U., Watanabe S., Hilschmann N.;
 RT "Rule of antibody structure. The primary structure of a monoclonal
 RT immunoglobulin L-chain of kappa-type, subgroup 3 (Bence-Jones protein
 RT T1). IV. The complete amino acid sequence and its significance for
 RT the mechanism of antibody production.";
 RL Hoppe-Seyler's Z. Physiol. Chem. 353:189-208 (1972).
 [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=81042304; Pubmed=6775818;
 RA Hietzer P.A., Max E.E., Seidman J.G., Maizel J.V. Jr., Leder P.;
 RT "Cloned human and mouse kappa immunoglobulin constant and J region
 RT genes conserve homology in functional segments.";
 RL Cell 22:197-207 (1980).
 [5]
 RP SEQUENCE (BENCE-JONES PROTEIN ROY).
 RA Hilschmann N., Barnikol H.U., Hees M., Langer B., Ponstingl H.,
 RL Steinmetz-Kayne M., Suter L., Watanabe S.;
 RL (in) Franck F., Shugar D. (eds.);
 RL Gamma globulins: structure and function, pp.57-74, Academic Press,
 RL New York (1969).
 [6]
 RP SEQUENCE (BENCE-JONES PROTEIN CUM).
 RX MEDLINE=68242259; Pubmed=5586923;
 RA Hilschmann N.;
 RT "The complete amino acid sequence of Bence Jones protein Cum (kappa-
 RT type)".;
 RL Hoppe-Seyler's Z. Physiol. Chem. 348:1718-1722 (1967).
 [7]
 RP SEQUENCE (BENCE-JONES PROTEIN AG).
 RX MEDLINE=69234734; Pubmed=4893682;
 RA Tilani K., Shioda T., Putnam F.W.;
 RT "The amino acid sequence of a kappa type Bence-Jones protein. 3. The
 RT complete sequence and the location of the disulfide bridges.";
 RL J. Biol. Chem. 244:3550-3560 (1969).
 [8]
 RP SEQUENCE (WALDENSTROM'S MACROGLOBULIN OU).
 RX MEDLINE=70201507; Pubmed=5447531;
 RA Kohler H., Shimizu A., Paul C., Putnam F.W.;
 RT "Macroglobulin structure: variable sequence of light and heavy
 RT chains.";
 RL Science 169:56-59 (1970).
 [9]
 RP SEQUENCE OF 1-33; 38-41 AND 62-80.
 RC TISSUE=Abdominal adipose tissue;
 RX MEDLINE=98249779; Pubmed=9588180;
 RA Olsen K.E., Sletten K., Westermark P.;
 RT "Extended analysis of AL-amyloid protein from abdominal wall
 RT subcutaneous fat biopsy: kappa IV immunoglobulin light chain.";
 RL Biochem. Biophys. Res. Commun. 245:713-716 (1998).
 [10]
 RP MISCELLANEOUS: The EU sequence has the INV (3) alloypic marker,
 CC 45-Ala and 83-Val. The ROI sequence has the INV (1,2) alloypic
 CC marker, 45-Ala and 83-Leu.
 [11]
 RP SMILARITY: Contains 1 immunoglobulin-like domain.
 [12]
 RP SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announce/>
 CC or send an email to license@ebi.ac.uk).
 [13]
 RP EMBL: J00241; AAA58989.1; -;
 DR EMBL: V00557; CAA23823.1; -;
 DR PIR: B90562; K3HU.
 DR PDB: 1DSB; 03-FEB-00.
 DR PDB: 1DS1; 09-FEB-00.
 DR PDB: 1D6V; 04-OCT-00.
 DR PDB: 1HEZ; 18-JUL-03.
 DR PDB: 1HKU; 12-MAR-97.
 DR PDB: 1I7Z; 08-AUG-01.
 DR PDB: 1MIM; 15-MAY-97.
 DR Genew; HGNC:5716; IGKC.

DR MIN; 147200; -;
 DR GO: GO:0003823; F:antigen binding; NAS.
 DR GO: GO:0006955; P:immune response; NAS.
 DR InterPro: IPR007110; Ig-like.
 DR InterPro: IPR003597; Ig-cl.
 DR InterPro: IPR003006; Ig_MHC.
 DR SMART: SM00047; Ig; 1.
 DR SMART: SM00407; Igcl; 1.
 DR PROSITE: PS00835; IG_LIKE; 1.
 DR PROSITE: PS00290; IG_MHC; 1.
 KW Immunoglobulin domain; Immunoglobulin C region; 3D-structure.
 FT NON TER 1 102 IG-LIKE.
 FT DOMAIN 5 102
 FT DISULFID 26 86
 FT DISULFID 106 106 INTERCHAIN (WITH A HEAVY CHAIN).
 FT VARIANT 83 83 V->L (IN INV(1,2) MARKER).
 FT CONFLICT 14 14 /FTID=VAR_003897.
 FT CONFLICT 57 57 D->N (IN REF. 7 AND 8).
 FT CONFLICT 57 57 E->Q (IN REF. 5 AND 6).
 SQ SEQUENCE 106 AA; 11609 MW; 51984D1FD372CE8 CRC64;
 Query Match 4.3%; Score 148; DB 1; Length 106;
 Best Local Similarity 29.5%; Pred. No. 0.004;
 Matches 28; Conservative 24; Mismatches 41; Indels 2; Gaps 1;
 QY 525 PoyvTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEMSNGL--QENNYKKTPTPLDDSGS 582
 DB 5 PSVTFPPSPDQLGSGTASVCLINNFYPRAKQWKVDNALQSGNSQESVTEGDSKDSST 64
 QY 583 FFYLSKLTVDKSRWQGNVFCSCVYHBAHNNHYTQ 617
 DB 65 YSLSTLTLLKADYKKYVACVETHGGLSSPYTK 99
 RESULT 85
 VCALL MOUSE STANDARD; PRT; 729 AA.
 AC P29533;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Vascular cell adhesion protein 1 precursor (V-CAM 1).
 GN VCAM1 OR VCAM-1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC STRAIN=FVB; TISSUE=Lung;
 RX MEDLINE=92181437; Pubmed=1371918;
 RA Heesoon C., Moy P., Tizard R., Chisholm P., Williams C., Wysk M.,
 RA Burkly L., Miyake K., Kincaid P., Lobb R.;
 RT "Cloning of murine and rat vascular cell adhesion molecule-1.";
 RL Biochem. Biophys. Res. Commun. 183:163-169 (1992).
 [2]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Lymph node;
 RX MEDLINE=93246254; Pubmed=7683304;
 RA Araki M., Araki K., Vassalli P.;
 RT "Cloning and sequencing of mouse VCAM-1 cDNA.";
 RL Gene 126:261-264 (1993).
 [3]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC STRAIN=129; TISSUE=Embryo;
 RX MEDLINE=94117008; Pubmed=7507076;
 RA Cybulsky M.I., Allan-Motamed M., Collins T.;
 RT "Structure of the murine VCAM1 gene.";
 RL Genomics 18:387-391 (1993).
 [4]
 RP SEQUENCE OF 1-693 FROM N.A. (ISOFORM 1).
 RC STRAIN=NIH Swiss, and 129/Sv;
 RA Kumar A.G., Dai Y.X., Kozak C.A., Mims M.P., Gotto A.M. Jr.,


```

Db 271 SGNATLTLIIMRMEDSGVYVEGVNLIGRDAEVELVQEKPFIVDISPGSQAUVGDS 330
Oy 139 LTLTLESPGSSPSVQCRSPGKNIG-----GKTLVSQLELDGSGTWTCTYLNQX 191
Db 331 VLTICAAIGCDSPSPSMRTQDSDPLANGVYRNEGAKSTIVLVSQFEDSHLCAVTCIOR 390
Oy 192 KVEFKIDIVLAFOKASSIVYKKEGQVEFSFPLA-----FTVEKL 232
Db 391 TLEKRTQVEVYSF-----PEDPVIKMSGPLVHGRPTVNCVTPVNVYFPHLEIELL 441
Oy 233 TGSGELMWQAEARASSKSMI-TFDLKNKEVSVKRYTQPKQ-MGKL-----PLHLTPQ 286
Db 442 KG-----ETTLMKKYFLBEMGKSLKLETLFTPLPTIEDGKSLVCLARLHSGEME 493
Oy 287 ALPOYASGNTLTLAEATGKLEHVNLVNRATQLOK-----NLTCVWGPTSPKMLML 342
Db 494 SEPKROSVQPLVYVAVP-----KTTTWSPSPLELGGSPVNLCSGDPAPRILMSR 548
Oy 343 KLENKAKVSKREKFWVJLN--PEAGNMOC-LSDSG--QVLESNIKVLPTWSTPVC 396
Db 549 QLNNGEIQPLSEBNTLTFMSYTKRDSGIVYCEGINEAGISKRSVELLIQV----- 598
Oy 397 PAPEKSCDKTHTCPELLIGSPVLPFP--KPKDTLMISRT---PEVTCV 442
Db 599 --SPKDIQLT-----VFPSKYVEGDTVIISCTCGNVFETWIL 635

RESULT 86
LAC1_RAT STANDARD; PRT; 104 AA.
ID LAC1_RAT
AC P20766;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE 19_Lambda-1 chain C region.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87305594; PubMed=3114047;
RA Steen M.L., Hellman L., Petersson U.;
RT "The immunoglobulin lambda locus in rat consists of two C lambda
RT genes and a single V lambda gene.";
RL Gene 55:75-84(1987).
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; M22520; AAA41419.1; ALT_INIT.
DR HSSP; P01842; 7EAB.
DR InterPro; IPR007110; IG-1-like.
DR InterPro; IPR003597; IG_C1.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; Ig_1.
DR SMART; SM00407; IgC1.1.
DR PROSITE; PS50835; IG_LIKE; 1.
DR PROSITE; PS00290; IG_MHC; 1.
KW Immunoglobulin domain; Immunoglobulin C region.
FT NON TER 1
FT DOMAIN 1
FT DISULFD 6 99 IG-LIKE.
FT DISULFD 27 85
FT DISULFD 103 103 INTERCHAIN (WITH HEAVY CHAIN).
SQ SEQUENCE 104 AA; 11565 MW; C8F7811F4BC878A CRC64;
Query Match 4.3%; Score 146; DB 1; Length 104;

```

```

Best Local Similarity 33.0%; Pred. No. 0.0052;
Matches 34; Conservative 19; Mismatches 44; Indels 6; Gaps 4;
Oy 521 QPR-EPQVYTLPRSDRLTNQVSLTCLVNGFSPSDIAVMESNGP-ENNYKTPPVLD 578
Db 1 QPKATPSVTLFPPSESLKDKATLVCMYDVPFGVATVWKADGITGQVETTP-FK 59
Oy 579 SDGSFPLYSKLTVDKSRWQGNVFCVMEALHNHYTKSLS 621
Db 60 QNKRYATSYLLTAKAMETHSNVSCVYTH---ENVYKSLS 99

RESULT 87
SRB2_HUMAN STANDARD; PRT; 387 AA.
ID SRB2_HUMAN
AC Q9P1R8; Q9WMA5; Q9NOK8.
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Signal-regulatory protein beta-2 precursor (SIRP-beta-2) (SIRP-b2).
GN SIRPB2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Placenta;
RX MEDLINE=21036165; PubMed=11185750;
RA Ichigotani Y., Matsumura S., Machida K., Oshima K., Iwamoto T.,
RA Yamaki K., Hayakawa T., Hamaguchi M.;
RT "Molecular cloning of a novel human gene (SIRP-B2) which encodes a new
RT member of the SIRP/SHS-1 protein family.";
RL J. Hum. Genet. 45:378-382(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21638749; PubMed=11780052;
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagusley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhand P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Frazer A.A., French L., Garner P.,
RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA Hammon S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Leharvalaho M.H., Leverisha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McDonachie L.J., McKay K., McMurray A.A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillips B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsey H.,
RA Rice C.M., Ross M.T., Scott C.E., Senta H.K., Showgreen R., Sims S.,
RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sutcliffe J.E.,
RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA Whitehead S.L., Whitlaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilmington L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.;
RT "The DNA sequence and comparative analysis of human chromosome 20.";
RL Nature 414:865-871(2001).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 3).
RC TISSUE=Lung;
RX MEDLINE=22388257; PubMed=12477932;
RA Straubeberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenner C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.T., Wang J., Heich F.,
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

```

RA Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahay J., Helton E., Keteeman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W.,
 RA Butlerfield Y.S.N., Krzywinski M.I., Skaleka U., Smalins D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -1- FUNCTION: Probable immunoglobulin-like cell surface receptor.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=3;
 CC Name=1;
 CC IsoId=Q9P1W8-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q9P1W8-2; Sequence=VSP_007027;
 CC Note=No experimental confirmation available;
 CC Name=3;
 CC IsoId=Q9P1W8-3; Sequence=VSP_007028;
 CC Note=No experimental confirmation available;
 CC -1- TISSUE SPECIFICITY: Detected in liver, and at very low levels in
 CC brain, heart, lung, pancreas, kidney, placenta and skeletal
 CC muscle.
 CC -1- SIMILARITY: Contains 2 immunoglobulin-like C1-type domain.
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AB042624; BAA95692.1; -;
 DR EMBL: AL138804; CAC00474.1; -;
 DR EMBL: BC020629; AAH20629.1; ALT_INT.
 DR Genew; HGNC:15757; SIRPB2.
 DR MIM; 605466; -;
 DR GO; GO:0007267; P:cell-cell signaling; TAS.
 DR GO; GO:0007242; P:intracellular signaling cascade; TAS.
 DR GO; GO:0008285; P:negative regulation of cell proliferation; TAS.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003597; Ig-cl.
 DR InterPro; IPR003506; Ig_v.
 DR InterPro; IPR003596; Ig_v.
 DR Pfam; PF00047; Ig_3.
 DR SMART; SM00407; IGcl; 2.
 DR SMART; SM00406; IGv; 1.
 DR PROSITE; PSS0835; IG_LIKE; 3.
 DR PROSITE; PS00290; IG_MHC; FALSE_NEG.
 DR Repeat; Signal; Transmembrane; Immunoglobulin domain; Glycoprotein;
 KM Alternative splicing.
 FT SIGNAL 1
 FT CHAIN 28
 FT DOMAIN 29 387
 FT TRANSMEM 29 360
 FT TRANSMEM 361 383
 FT DOMAIN 384 387
 FT DOMAIN 29 137
 FT DOMAIN 146 245
 FT DOMAIN 252 340
 FT DISULFID 53 119
 FT DISULFID 168 226
 FT DISULFID 271 329
 FT CARBOHYD 243 243
 FT CARBOHYD 268 268

FT CARBOHYD 309 309 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 317 317 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPPLIC 1 33 Missing (in isoform 2).
 FT VARSPPLIC 144 360 /FTid=VSP_007027.
 FT VARSPPLIC 144 360 Missing (in isoform 3).
 FT CONFLICT 263 263 /FTid=VSP_007028.
 FT CONFLICT 286 286 A -> V (IN REF. 2).
 FT CONFLICT 286 286 L -> S (IN REF. 2).
 SQ SEQUENCE 387 AA, 42495 MW, F7F20C9F8E0E64B CRC64;
 Query Match 4.3%; Score 146; DB 1; Length 387;
 Best Local Similarity 24.8%; Pred. No. 0.029;
 Matches 57; Conservative 37; Mismatches 114; Indels 22; Gaps 8;
 QY 412 ELGSPSV-FLPPPKXTLMIS---RTPETCVVDVSH--EDPEVKENNYVGVGVH 464
 DB 131 EFKSGGTEMALGAPSPVIVGPAARTTPERTVTSFCSHGPSPPDITLKWFKNGNELS 190
 QY 465 NAKTPREQVNSTRVVSVLTVLHODMLNGEKVKSNKALPA-PIEKT--ISKAKQ 521
 DB 191 DQTNVDPPGQSVAYSINSTARVVDLPMDVRSQVCEVAHVTLQGDPLAGTANLSALIV 250
 QY 522 PREPVYTLPPSRDELTKNQVSLTCLVKGYFSPDAVEMWESNGQDENNYKTPPVLDSDG 581
 DB 251 PPTLEV-TQQPMR---AGNQVNVTCQVRKFYQSIQLTFLENGVCQRETAFTLENKDG 306
 QY 582 SEFLYSKLTVDYSRMOQGVNFSQVMEALNHHYTKSLSPGLQDTE 631
 DB 307 TYNMTSWFLVNIISDQRDVVLTCQVKHDG-----QLAVSKRLALEVT 348
 RESULT 88
 LAC_CHICK
 ID LAC_CHICK STANDARD; PRT; 103 AA.
 AC P20763;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Ig lambda chain C region.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_Taxid=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87218480; PubMed=3107981;
 RA Parvati R., Ziv E., Lenter F., Tel-Or S., Burshtein Y., Schechter I.,
 RT "Analyses of chicken immunoglobulin light chain cDNA clones indicate
 RT a few germline V lambda genes and allotypes of the C lambda locus";
 RL EMBO J. 6:97-102(1987).
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X04768; CAA28461.1; -;
 DR PIR; B26167; B26167.
 DR HSSP; P01842; 7FAB.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003597; Ig-cl.
 DR InterPro; IPR003506; Ig_MHC.
 DR Pfam; PF00047; Ig_1.
 DR SMART; SM00407; IGcl; 1.
 DR PROSITE; PSS0835; IG_LIKE; 1.
 DR PROSITE; PS00290; IG_MHC; 1.
 KM Immunoglobulin domain; Immunoglobulin C region.
 FT NON_TER 1 1

```

FT DOMAIN 6 99 IG-LIKE.
FT DISULFID 28 85 INTERCHAIN (WITH HEAVY CHAIN).
FT DISULFID 103 103 N -> D.
FT VARIANT 90 90
SO SEQUENCE 103 AA; 11361 MW; 77BF341B511B91B2 CRC64;

Query Match 4.3%; Score 145.5; DB 1; Length 103;
Best Local Similarity 32.2%; Pred. No. 0.0056;
Matches 29; Conservative 19; Mismatches 39; Indels 3; Gaps 3;

Oy 521 QPR-EPQVYTLPPSRDELTK-NQVSLTCLVKGFPSDIAVEMESNGOPENNYKTPPYLD 578
Db 1 QPKVAPRTTLPPSRKEELNEATKATLVCLINDFYPSPTVDMVIDGSTRSG-ETTAPORO 59
Oy 579 SDGSFPLYSKLTVDKSRMOGQNVFSCSVME 608
Db 60 SNSQYMASSYLSLSASDMSHETVCRAVTH 89

RESULT 89
KACB_RAT STANDARD; PRT; 106 AA.
ID _KACB_RAT
AC P01835;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig kappa chain C region, B allele.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Jouvalin;
RX MEDLINE=82082587; PubMed=62733908;
RA Sheppard H.W., Gutman G.A.;
RT "Allelic forms of rat kappa chain genes: evidence for strong
RL selection at the level of nucleotide sequence";
RL Proc. Natl. Acad. Sci. U.S.A. 78:7064-7068(1981).
RN [2]
RP SEQUENCE (BENCE-JONES PROTEIN S211).
RX MEDLINE=7521238; PubMed=807630;
RA Starace V., Querintan P.;
RT "The primary structure of a rat kappa Bence Jones protein:
RL phylogenetic relationships of V- and C-region genes.";
CC -1 SIMILARITY: Contains 1 immunoglobulin-like domain.
DR HSP; A93901; KIRTB.
DR HSP; P01842; 2MCG.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003597; Ig_c1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig_1.
DR SMART; SM00407; IGc1; 1.
DR PROSITE; PS00835; IG_LIKE; 1.
DR PROSITE; PS00290; IG_MHC; 1.
KM Immunoglobulin domain; Immunoglobulin C region.
FT NON_TER 1 1
FT DOMAIN 5 102 IG-LIKE.
FT DISULFID 26 86
FT DISULFID 106 106 INTERCHAIN (WITH A HEAVY CHAIN).
FT CONFLICT 2 2 D -> N (IN REF. 2).
FT CONFLICT 30 30 N -> K (IN REF. 2).
FT CONFLICT 48 48 MISSING (IN REF. 2).
FT CONFLICT 79 79 E -> Q (IN REF. 2).
FT CONFLICT 87 87 E -> Q (IN REF. 2).
FT CONFLICT 98 98 V -> VV (IN REF. 2).
FT CONFLICT 100 100 S -> N (IN REF. 2).
SO SEQUENCE 106 AA; 11601 MW; 4CFA7CA820D1CA36 CRC64;

Query Match 4.2%; Score 145; DB 1; Length 106;
Best Local Similarity 31.5%; Pred. No. 0.0062;
Matches 29; Conservative 22; Mismatches 29; Indels 12; Gaps 2;

```

```

Oy 525 POVYTLPPSRDELTKNQVSLTCLVKGFPSDIAVEMESNGOPENNYKTPPYLD 579
Db 5 PTYSIFPSTBQTLATGASVYCLMANNYPKIDISVKMIDGTERRD-----GVLDSTVDQ 59
Oy 580 --DGSPFLYSKLTVDKSRMOGQNVFSCSVME 609
Db 60 SKDSTYSMSSTLSLTAKDYESHNLVTCVYVHK 91

RESULT 90
HB2C_PIG STANDARD; PRT; 261 AA.
ID _HB2C_PIG
AC P15982;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUN-1999 (Rel. 38, Last annotation update)
DE SLA class II histocompatibility antigen, DQ haplotype C beta chain
DE precursor.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxId=9823;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90361905; PubMed=2391424;
RA Gustafsson K., Leguenn C., Hirsch F., Germana S., Pratt K.,
RA Sachs D.H.;
RT "Class II genes of miniature swine. IV. Characterization and
RL expression of two allelic class II DQB cDNA clones.";
RL J. Immunol. 145:1946-1951(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; M31497; AAA31084.1; -
DR EMBL; M32117; AAA53110.1; -
DR PIR; A60404; A60404.
DR HSP; P13760; 2SEB.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003597; Ig_c1.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR000353; MHC_II_beta.
DR Pfam; PF00047; Ig_1.
DR Pfam; PF00969; MHC_II_beta; 1.
DR PRODOM; PD000328; MHC_II_beta; 1.
DR SMART; SM00407; IGc1; 1.
DR PROSITE; PS00835; IG_LIKE; 1.
DR PROSITE; PS00290; IG_MHC; 1.
KM MHC II; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 31
FT CHAIN 32 261
FT DOMAIN 32 126
FT DOMAIN 127 220 DO HAPLOTYPE C BETA CHAIN.
FT DOMAIN 221 230 EXTRACELLULAR BETA-2.
FT TRANSMEM 231 251 CONNECTING PEPTIDE.
FT DOMAIN 252 261 CYTOPLASMIC TAIL.
FT DISULFID 47 111 BY SIMILARITY.
FT DISULFID 149 205 BY SIMILARITY.
FT CARBOHYD 51 51 N-LINKED (GLCNAC...) (POTENTIAL).
SO SEQUENCE 261 AA; 29550 MW; AA9581F2A3B1969D CRC64;

Query Match 4.2%; Score 145; DB 1; Length 261;
Best Local Similarity 24.7%; Pred. No. 0.02;
Matches 49; Conservative 36; Mismatches 93; Indels 20; Gaps 6;
Oy 417 PSVTLPPKPKDRLMISRTPEVLCVAVDVSHEDPEVFNMYDVGVHNAKTPREQYN 476

```

```

DB 36 PDQFVQFK-GECEYFNGTORAGVARIYNGEHLRFSDVGEFAATVPLGRPEADSN 94
QY 477 STYRVSVLTVLHQMNLNGEKYCKVSNKALPAPIEKITSKAGQREPOVYTLPPSRDE 536
DB 95 SQDVLQWRA-EVDVRCCKRNYQIE-----EGTTLQRRVQ-----TVTISPSKAE 139
QY 537 LTNGQVSLCTLVGFFPSDIAVWESNGQPEN-NYKTTPLVDSDFSFLXSLTLDKSR 595
DB 140 ALMHNHLVCAVTDFFPSQVKVQMFRRNGQETAGVSTPLIRNGD---WTYQVLMLENN 196
QY 596 MOQNFVSCSVMEALHN 613
DB 197 LQRGDVYTCRVEHSION 214

RESULT 91
SHS1_RAT STANDARD; PRT; 509 AA.
ID SHS1_RAT 008951, 070426, 09QW15;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE Protein-tyrosine phosphatase non-receptor type substrate 1 precursor
DE (SHP substrate-1) (SHPS-1) (inhibitory receptor SHPS-1) (Signal-
DE regulatory protein alpha-1) (Sirt-alpha-1) (Brain Ig-like molecule
DE with tyrosine-based activation motifs) (Bit) (Macrophage fusion
DE receptor) (Macrophage membrane protein MFP150).
GN PRPN1 OR SHPS1 OR SIRT OR BIT OR MFR.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
[1]
SEQUENCE FROM N.A., SEQUENCE OF 42-60; 68-91; 128-137; 150-158;
RP 174-189; 192-202; 204-212; 218-237; 259-270; 279-282; 405-415 AND
RP 446-453. N-GLYCOSYLATION, PHOSPHORYLATION ON TYROSINE RESIDUES, AND
RP INTERACTIONS WITH PTPN6 AND PTPN11.
RC TISSUE=Fetal Fibroblast;
RX MEDLINE=97098667; PubMed=8943344;
RA Fujioaka Y., Matozaki T., Noguchi T., Iwamatsu A., Yamao T.,
RA Takahashi N., Tsuda M., Takada T., Kasuga M.;
RT "A novel membrane glycoprotein, SHPS-1, that binds the SH2-domain-
RT containing protein tyrosine phosphatase SHP-2 in response to mitogens
RT and cell adhesion.";
RN Mol. Cell. Biol. 16:6887-6899(1996).
[2]
SEQUENCE FROM N.A., SEQUENCE OF 32-48 AND 446-453, FUNCTION, AND
RP PHOSPHORYLATION ON TYROSINE RESIDUES.
RC STRAIN=Sprague-Dawley; TISSUE=Brain;
RX MEDLINE=97415431; PubMed=9271230;
RA Sano S.-I., Ohnishi H., Omori A., Hasegawa J., Kubota M.;
RT "BIT, an immune antigen receptor-like molecule in the brain.";
RN FEBS Lett. 411:327-334(1997).
[3]
SEQUENCE FROM N.A., SEQUENCE OF 99-107; 128-149; 192-217; 405-417;
RP 419-429; 446-467 AND 496-506. N-GLYCOSYLATION, AND TISSUE SPECIFICITY.
RC STRAIN=Fischer 344; TISSUE=Macrophage;
RX MEDLINE=98449911; PubMed=9774638;
RA Saginario C., Sterling H., Beckers C., Kobayashi R., Solimena M.,
RA Ulu E., Vignery A.;
RT "MFR, a putative receptor mediating the fusion of macrophages.";
RN Mol. Cell. Biol. 18:6213-6223(1998).
[4]
SEQUENCE OF 1-419 FROM N.A., AND TISSUE SPECIFICITY.
RC STRAIN=MG/Rij; TISSUE=Alveolar macrophage;
RX MEDLINE=98375871; PubMed=9712053;
RA Adams S., van der Laan L.J.W., Vernon-Wilson E.,
RA Renardel de Lavalette C., Doepf E.A., Dijkstra C.D., Simmons D.L.,
RT "Signal-regulatory protein is selectively expressed by myeloid and
RT neuronal cells.";
RN J. Immunol. 161:1853-1859(1998).

```

```

RN [5]
RP PHOSPHORYLATION IN RESPONSE TO EGF, AND INTERACTION WITH PTPN11.
RX MEDLINE=98008865; PubMed=9344856;
RA Ochi F., Matozaki T., Noguchi T., Fujioaka Y., Yamao T., Takada T.,
RA Tsuda M., Takada H., Fukunaga K., Okabayashi Y., Kasuga M.;
RT "Epidermal growth factor stimulates the tyrosine phosphorylation of
RT SHPS-1 and association of SHPS-1 with SHP-2, a SH2 domain-containing
RT protein tyrosine phosphatase.";
RN Biochem. Biophys. Res. Commun. 239:483-487(1997).
[6]
RP PHOSPHORYLATION OF TYR-477 AND TYR-501, AND MUTAGENESIS OF TYR-436;
RX TYR-460; TYR-477 AND TYR-501.
RX MEDLINE=98204923; PubMed=9535915;
RA Takada T., Matozaki T., Takada H., Fukunaga K., Noguchi T.,
RA Fujioaka Y., Okazaki I., Tsuda M., Yamao T., Ochi F., Kasuga M.;
RT "Roles of the complex formation of SHPS-1 with SHP-2 in
RT insulin-stimulated mitogen-activated protein kinase activation.";
RN J. Biol. Chem. 273:9234-9242(1998).
CC -1- FUNCTION: Immunoglobulin-like cell surface receptor for CD47. Acts
CC as docking protein and induces translocation of PRPN6,
CC PTPN11 and other binding partners from the cytosol to the
CC plasma membrane. Supports adhesion of cerebellar neurons, neurite
CC outgrowth and glial cell attachment. May play a key role in
CC intracellular signaling during synaptogenesis and in synaptic
CC function. Involved in the negative regulation of receptor tyrosine
CC kinase-coupled cellular responses induced by cell adhesion, growth
CC factors or insulin. Mediates negative regulation of phagocytosis,
CC mast cell activation and dendritic cell activation. CD47 binding
CC prevents maturation of immature dendritic cells and inhibits
CC cytokine production by mature dendritic cells. May play a role in
CC the release of nitric oxide by macrophages (By similarity).
CC -1- SUBUNIT: Binds PTPN11 when tyrosine-phosphorylated, except in
CC macrophages, where it primarily binds PTPN6. Binds GRB2 in vitro.
CC Binds FGR. Binds JAK2 irrespective of its phosphorylation status
CC and forms a stable complex. Binds SCAPI and/or SCAPI2. The
CC resulting complex recruits Fyb. Binds PRK2B (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: Highly expressed in brain, spleen, lung, liver
CC and kidney. Detected at lower levels in heart. Highly expressed in
CC alveolar and peritoneal macrophages, and at lower levels in
CC dendritic cells.
CC -1- PTM: N-glycosylated.
CC -1- PTM: Phosphorylated on tyrosine residues in response to insulin,
CC cell adhesion or epidermal growth factors. Dephosphorylated by
CC PTPN11.
CC -1- SIMILARITY: Contains 2 immunoglobulin-like C1-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL, D85133; BAAL2734.1; -
DR EMBL, D38468; BAA20368.1; -
DR EMBL, U62328; AAC68478.1; -
DR EMBL, AF055065; AAC18089.1; -
DR HSSP, P01703; 7FAB.
DR InterPro: IPR007110; Ig-like.
DR InterPro: IPR003587; Ig_C1.
DR InterPro: IPR003006; Ig_MHC.
DR Pfam, PF00047; Ig1; 3.
DR SMART, SM00407; IG1; 2.
DR PROSITE, PS00290; IG_LIKE; 3.
DR Repeat, Signal; Transmembrane, Immunoglobulin domain; SH3-binding;
KW Glycoprotein; Phosphorylation.
FT SIGNAL 1 31
FT CHAIN 32 509
FT TYPE SUBSTRATE 1.

```

```

FT DOMAIN 28 373 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 374 394 POTENTIAL.
FT CYTOPLASMIC 395 509 POTENTIAL.
FT DOMAIN 32 138 IG-LIKE V-TYPE.
FT DOMAIN 150 248 IG-LIKE C1-TYPE 1.
FT DOMAIN 255 349 IG-LIKE C1-TYPE 2.
FT DISULFID 55 122 POTENTIAL.
FT DISULFID 172 229 POTENTIAL.
FT DISULFID 274 332 POTENTIAL.
FT SITE 436 439 SH2-BINDING (POTENTIAL).
FT SITE 446 451 SH3-BINDING (POTENTIAL).
FT SITE 460 463 SH2-BINDING (POTENTIAL).
FT SITE 477 480 SH2-BINDING (POTENTIAL).
FT SITE 501 504 SH2-BINDING (POTENTIAL).
FT MOD_RES 436 436 PHOSPHORYLATION (BY TYR-KINASES) (POTENTIAL).
FT MOD_RES 460 460 PHOSPHORYLATION (BY TYR-KINASES) (POTENTIAL).
FT MOD_RES 477 477 PHOSPHORYLATION (BY TYR-KINASES) (POTENTIAL).
FT MOD_RES 501 501 PHOSPHORYLATION (BY TYR-KINASES) (POTENTIAL).
FT CARBOHYD 54 54 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 93 93 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 169 169 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 181 181 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 205 205 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 242 242 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 246 246 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 271 271 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 312 312 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 320 320 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 345 345 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 436 436 Y->F: ABOLISHES TYROSINE PHOSPHORYLATION AND PTPN11 BINDING; WHEN ASSOCIATED WITH F-436, F-477 AND F-501.
FT MUTAGEN 460 460 Y->F: ABOLISHES TYROSINE PHOSPHORYLATION AND PTPN11 BINDING; WHEN ASSOCIATED WITH F-436, F-477 AND F-501.
FT MUTAGEN 477 477 Y->F: STRONGLY REDUCES INSULIN-INDUCED TYROSINE PHOSPHORYLATION AND PTPN11 BINDING. ABOLISHES TYROSINE PHOSPHORYLATION AND PTPN11 BINDING. WHEN ASSOCIATED WITH F-436, F-460 AND F-501.
FT MUTAGEN 501 501 Y->F: STRONGLY REDUCES INSULIN-INDUCED TYROSINE PHOSPHORYLATION AND PTPN11 BINDING. ABOLISHES TYROSINE PHOSPHORYLATION AND PTPN11 BINDING; WHEN ASSOCIATED WITH F-436, F-460 AND F-477.
FT CONFLICT 8 8 MISSING (IN REF. 4).
FT CONFLICT 10 10 F->I (IN REF. 3).
FT CONFLICT 25 25 S->C (IN REF. 4).
FT CONFLICT 58 58 KR->MP (IN REF. 3; AA SEQUENCE).
FT CONFLICT 99 100 G->A (IN REF. 2).
FT CONFLICT 162 189 D->N (IN REF. 3).
FT CONFLICT 189 189 N->L (IN REF. 3; AA SEQUENCE).
FT CONFLICT 205 205 N->G (IN REF. 3; AA SEQUENCE).
FT CONFLICT 209 205 N->F (IN REF. 3; AA SEQUENCE).
FT CONFLICT 405 405 E->P (IN REF. 3; AA SEQUENCE).
FT CONFLICT 416 421 NARE->EGON (IN REF. 3; AA SEQUENCE).
FT CONFLICT 418 421 R->E (IN REF. 3; AA SEQUENCE).
FT CONFLICT 450 450 MISSING (IN REF. 3; AA SEQUENCE).
FT CONFLICT 499 499 MISSING (IN REF. 3; AA SEQUENCE).
SQ SEQUENCE 509 AA: 55690 MW: 581FE0A4D429F4 CRC64;
Query Match 4.24; Score 145; DB 1; Length 509;
Best Local Similarity 21.54; Pred. No. 0.048;
Matches 82; Conservative 64; Mismatches 161; Indels 74; Gaps 16;

```

```

Qy 314 LVWRATQLOKNTLCEWGPSTSPKMLSLKENKAKVSKREKRPVWLVNBEAGMOCILS 373
Db 92 TNVSDATK-RNNMD-----FSICISN-----VTPEDAGTYCVKF 125
Qy 374 DSGVVLIESNIK-----VLPWSTVPVPCPAPEPKSCDKHTCPBLGGPSVFLPPPKKD 428
Db 126 QKGIIVEPDTREIKSGGGTTLVYLAKP---SSPEVSG-----PDSRGSF----- 164
Qy 429 TLMISREVTGVVDVSHEDPEYKFWYVDGVVNAKTRPEEGVNSRYRVSVLTV- 487
Db 165 ----GQVNFCTCSYSGSPNN--ITLKMKGKELSHLETTISKS-IVSYNISTVSVK 217
Qy 488 LHQDWLNGKYEKCVSKNALPA-PIEKTISKAKGQPREPOVYTLPSRDELTL-KNGVSLT 545
Db 218 LSPEDHSRIY-CEVAHVLEGRPLNGTANFSNIIRSP---TLKITQGLTPASQVNL 273
Qy 546 CLVGFYPSDIAVWESNGQPENNYKTPPVLDGSGFFLYSKLTYDKSRWQGNVFSQS 605
Db 274 CQVQKFPKALQNLWLENGMLSRDKPEHFTDRDGYNTYSPLVNSAHRBDVVFYTCQ 333
Qy 606 VMHE---ALNNHTQKSLIS 623
Db 334 VERDSQPAITENHTVRAFAHS 354
RESULT 92
NCM2_MOUSE STANDARD; PRT; 837 AA.
ID NCM2_MOUSE
AC 035136; 035962;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Neural cell adhesion molecule 2 precursor (N-CAM 2) (RB-8 neural cell adhesion molecule) (RAMB12).
GN NCAM2 OR CCAM OR RNCAM.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).
RC STRAIN=BALB/c; TISSUE=Olfactory neuroepithelium;
RX MEDLINE=9736828; Pubmed=9221781;
RA Yoshihara Y., Kawasaki M., Tamada A., Fujita H., Hayashi H., Kagamiyama H., Mori K.;
RT "CCAM: A new member of the neural cell adhesion molecule family related to zone-to-zone projection of olfactory and vomeronasal axons.";
RT J. Neurosci. 17:5830-5842(1997).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM SHORT).
RC STRAIN=C57BL/6J; TISSUE=Olfactory epithelium;
RX MEDLINE=97476194; Pubmed=9331170;
RA Alenius M., Bohm S.;
RT "Identification of a novel neural cell adhesion molecule-related gene with a potential role in selective axonal projection.";
RT J. Biol. Chem. 272:26083-26086(1997).
CC -1- FUNCTION: May play important roles in selective fasciculation and zone-to-zone projection of the primary olfactory axons.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (long isoform) and attached to the membrane by a GPI-anchor (short isoform).
CC -1- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=2;
Name=Long;
Name=Short;
Name=IsoId=O35136-1; Sequence=Displayed;
Name=IsoId=O35136-2; Sequence=VSP_002590.
CC -1- TISSUE SPECIFICITY: Expressed in subsets of both olfactory and vomeronasal neurons in a zone-specific manner.
CC -1- SIMILARITY: Contains 5 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 2 fibronectin type III domains.
CC -----
This SWISS-Prot entry is copyright. It is produced through a collaboration

```

D	b		347	DKSPDGRILEVKGQHGSSLIHRDVKLSDSGRYCEAASRIQGHORSNHLIDIEAP-KFVS	405
O	y		207	ASSIVTYKEEBOVEFFSEPLAFPTVEKLGTSGELWMQAERASSSKSMITFDLKNKEYSKRV	266
		:	:	:	:
D	b		406	NQMTVMSEGNPINISCDVT-----ANRPASIMHRRKKL-----LPAXNT	446
O	y		267	TODPKLOMKKLLPLHLTPOLPOYAGSGLTLALEAKTGKHOFVNVLVMPRATOLOKXL	326
		:	:	:	:
D	b		447	THLKTHSVGKMILLEIA-PTSNDNF---GRYNCTATRTIGTRFOEYLLEADV-----	495
O	y		327	TCEWVGSTPSKMLSLKLENKAQVSRERP-----VVNLNPESAGMGOCLSD	374
		:	:	:	:
D	b		496	-----PSSPHGVKTIIELSQTTAKIS-FNKPESHGGVPRIHHYOVDVAEVASETKVIARSH	548
O	y		375	SGQVL-----LESN-----IKVLPTWSTPVPCPAPEPKSCDKTHT	409
		:	:	:	:
D	b		549	GVOGMVNVLSLEERTTYETIIRVAAVNGKGDDYSKIETFPQ--LPVNEPS-----	596
O	y		410	CPELLGSPVFLPPPKKDITLMISRPEVTCVVVDVSHEDPEVKFMNYVDGEVHNAAKTG	469
		:	:	:	:
D	b		597	-PSIHGPS-----SGKFSKISTFKQ-----DDGAPFILEY-----IVKYRSK	633
O	y		470	PREEQY-----NSTRYRVSVLTVLHQDLNKEVKKCSNKALPAPIEKITSKAKGP	522
		:	:	:	:
D	b		634	DKEOEWLEKKVQGNKHIIIE-----HLQWTMGYEVOITANRL-----GY	674
O	y		523	REPOVY--TLPP	532
		:	:	:	:
D	b		675	SEPVTVFYEFMSMP	686
 RESULT_93					
C	l66_CARAU	ID	C166_CARAU	STANDARD;	PRT; 555 AA.
A	C		Q90304;		
D	T		01-NOV-1997 (Rel. 35, Created)		
D	T		30-MAY-2000 (Rel. 39, Last sequence update)		
D	T		10-OCT-2003 (Rel. 42, Last annotation update)		
D	E		CD16 antigen homolog precursor (Neuroilin) (DM-GRASP homolog).		
O	S		Carassius auratus (Goldfish).		
O	C		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
O	C		Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;		
O	C		Cyprinidae; Carassius.		
O	X		NCBI_TaxId=7957;		
R	N	[1]	SEQUENCE FROM N.A.		
R	P		TISSUE=Retina;		
R	C		Laessing U.;		
R	L		Submitted (JUN-1999) to the EMBL/GenBank/DDB databases.		
R	N	[2]	SEQUENCE OF 30-552 FROM N.A., AND PARTIAL SEQUENCE.		
R	P		MEDLINE=94299040; PubMed=8026643;		
R	X		Laessing U., Giordano S., Stecher B., Lottspeich F., Stuermer C.A.O.;		
R	T		"Molecular characterization of fish neuroilin: a growth-associated		
R	T		cell surface protein and member of the immunoglobulin superfamily in		
R	T		the fish retinoectal eyetem with similarities to chick protein		
R	T		DM-GRASP/SC-1/BEN.";		
R	L		Differentiation 56:21-29(1994).		
C	C	-1-	FUNCTION: CELL ADHESION MOLECULE. INVOLVED IN NEURITE EXTENSION BY		
C	C	-1-	NEURONS VIA HETEROPHILIC AND HOMOPHILIC INTERACTIONS.		
C	C	-1-	SUBCELLULAR LOCATION: TYPE I membrane protei.		
C	C	-1-	TISSUE SPECIFICITY: PRESENT ON ALL RETINAL GANGLION CELLS (RGCS)		
C	C		AND THEIR AXONS (IN EMBRYO). ABSENT FROM MATURE AXONS ALONG MOST		
C	C		OF THEIR LENGTH, BUT IS PRESENT ON NEW AND GROWING AXONS DERIVED		
C	C		FROM THE RGCS AT THE RETINAL MARGIN. REMAINS ON ADULT RGCS ONLY AT		
C	C		CELL-CELL CONTACT SITES AND IS CONTINUOUSLY FOUND IN THE RETINAL		
C	C		AXON TERMINAL ARBOR LAYERS OF THE ADULT TECTUM.		
C	C	-1-	SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.		
C	C	-1-	SIMILARITY: Contains 2 immunoglobulin-like V-type domains.		
C	C	-1-	This SWISS-PROT entry is copyright. It is produced through a collaboration		
C	C		between the Swiss Institute of Bioinformatics and the EMBL Outstation -		

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announce/>
 CC or send an email to license@ebi.ac.uk).

CC EMBL; L25056; AAC38015.2; -
 DR PIR; I50478; I50478.
 DR HSSP; Q13740; 1KJC.
 DR InterPro; IPR007110; IG-like.
 DR InterPro; IPR003006; IG_MHC.
 DR InterPro; IPR003596; IG_V.
 DR Pfam; PF00047; Ig; 3.
 DR SMART; SM00406; IG; 1.
 DR PROSITE; PS00835; IG LIKE; 4.
 DR PROSITE; PS00290; IG_MHC; 1.
 DR Signal; Developmental protein; Cell adhesion; Immunoglobulin domain;
 KW Repeat; Glycoprotein; Transmembrane.
 FT SIGNAL 1 22 POTENTIAL.
 FT CHAIN 23 555 CD166 ANTIGEN HOMOLOG..
 FT DOMAIN 23 499 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 500 520 POTENTIAL.
 FT DOMAIN 521 555 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 31 127 IG-LIKE V-TYPE 1.
 FT DOMAIN 131 229 IG-LIKE V-TYPE 2.
 FT DOMAIN 323 397 IG-LIKE C2-TYPE 1.
 FT DOMAIN 319 397 IG-LIKE C2-TYPE 2.
 FT DOMAIN 406 484 IG-LIKE C2-TYPE 3.
 FT DISULFID 38 110 POTENTIAL.
 FT DISULFID 154 217 POTENTIAL.
 FT DISULFID 263 306 POTENTIAL.
 FT DISULFID 426 470 POTENTIAL.
 FT CARBOHYD 92 92 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 171 171 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 350 350 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 441 441 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 465 465 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 555 AA; 60371 MW; 544A8014F00BFF68 CRC64;

Query Match 4.2%; Score 144.5; DB 1; Length 555;
 Beest Local Similarity 19.1%; Pred. No. 0.058;
 Matches 118; Conservative 91; Mismatches 203; Indels 207; Gaps 30;

QY VLGGKSDVELTCTASOKS---IQFHKNQNOIKIIGNOSFLLTKGSK-----LNDK 79
 DB 26 VILGELTIVPCNDGTKKPDGLFTTKK---YKDDSPEDLLVKAQKDEATVSATDG 82
 QY ADSRRSLWDQGNFPLIKLIKIEDSDTYICEVEDQKEEVQLVFGLTANSPTHLQGS 139
 DB 83 YKSNVSI--ANSSLLIARGLADQVFTC-----MVVSFTN-----LEYSV 123
 QY 140 TLTLESPPGSSPVQCSPPKKNIOGCKTISVSOLELQDSGTWTCTVLQNKQVEFKIDI 199
 DB 124 EVKHKRP-SAPVIK--NNAKELENGKLTQLGECVENA----- 159
 QY 200 VVLAFOKASSIVYKKEGEV-----EFSFLATFVEKLTGSGELMWAERASSKSWTF 254
 DB 160 -----NPPADLIWKKNQTLVDDKTIITITSTTKDKTG-----LSSTSRLQY 204
 QY 255 DLKKNKSVKRVTDQPKLQMGKLLPLHLTPQALQVAGSGLTLALEAKTKGLH-----OE 311
 DB 205 TAKREDEVE-SQFTCAKIVMG---PDQVSEPSFP-----IHYFTEK 242
 QY 312 VNLVVMRATQLOK---NLTCGVWGPTSPKMLSLKENKAVSKREKPVVVL-----NP 363
 DB 243 VSLQVVSQSPIREGEDVTLKCOADGNPP--TSFENIKKQKVTATVDKDYTLTGVTBA 299
 QY 364 EAGWQCLLSDSGOVLLES-----NIKVLPTMSPPVPCAPRPSCKTKTTCPELL 414
 DB 300 DSGIYKCSLNDN--WVESTQFTVVSFLDVSFLTGTG----- 334
 QY 415 GGPSPVFLPPPKKDTLMISRTPEVTCVVDVSHEDPEVKFMVYDVGVVHNAKTKPREEO 474

DB 335 -----VLKNVGENLTVSLD-KNASSBAKVTWTKD-----NRKLDKLPDF 372
 QY 475 YNSTRYRVSVLTVLHQDWLNGKEKCKVKNAKLPAPKEKTSKAKGQPREPOVYTLPPSR 534
 DB 373 SKLTYSNAGL-----YCDVSTIEG-----IKRSLSELTVEGIPKITSLTXHR 415
 QY 535 DELTKNQVSLTCLVKGFYPSDIAVEWESNGQP-----NNYKTPPVLDSDGSEFVLSKLT 530
 DB 416 SSGDKKHKV-LTCEAEAGSKPRD--VQMSVNTQNTDNEVSYNNKGAT-----YKLT 459
 QY 591 VDKSRMQQGVFSGSVWME 609
 DB 460 VVPSK--NLTVSCLVTVNK 475

RESULT 94
 SHS1_HUMAN STANDARD; PRT; 503 AA.
 AC P78324; O00683; O43799; Q8N517; Q8TAL8; Q9H022; Q9UDX2; Q9UIJ6;
 AC O9Y4U9;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Protein-tyrosine phosphatase non-receptor type substrate 1 precursor
 DE (SHP substrate-1) (SHPs-1) (inhibitory receptor SHP-1) (Signal-
 DE regulatory protein alpha-1) (SHP-alpha-1) (SHP-alpha-2) (SHP-alpha-
 DE 3) (MYD-1 antigen) (Brain Ig-like molecule with tyrosine-based
 DE activation motifs) (Blt) (Macrophage fusion receptor) (p84).
 GN PTPNS1 OR SHS1 OR SIRP OR MYD1 OR BIT OR MFR.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Brain;
 RX MEDLINE=9722339; PubMed=9070220;
 RA Yamao T., Matozaki T., Amano K., Matsuda Y., Takahashi N., Ochi F.,
 RA Fujioka Y., Kasuga M.;
 RT "Mouse and human SHP-1: molecular cloning of cDNAs and chromosomal
 RT localization of genes.";
 RL Biochem. Biophys. Res. Commun. 231:61-67(1997).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 1), POLYMORPHISM, PHOSPHORYLATION,
 RP N-GLYCOSYLATION, AND INTERACTIONS WITH PTPNS1, PTPNS2 AND GRB2.
 RC TISSUE=Placenta;
 RX MEDLINE=97215901; PubMed=9062191;
 RA Karilainenkov A., Chen Z., Sures I., Wang H., Schilling J.,
 RA Ulrich A.;
 RT "A family of proteins that inhibit signalling through tyrosine kinase
 RT receptors.";
 RL Nature 386:181-186(1997).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 3), AND VARIANTS ASP-95; LEU-96; ASN-100;
 RP ARG-107; GLY-109 AND VAL-131.
 RC TISSUE=Monocytes;
 RX MEDLINE=98143722; PubMed=9485180;
 RA Brooke G.P., Parsons K.R., Howard C.J.;
 RT "Cloning of two members of the SIRP alpha family of protein tyrosine
 RT phosphatase binding proteins in cattle that are expressed on monocytes
 RT and a subpopulation of dendritic cells and which mediate binding to
 RT CD4 T cells.";
 RL Eur. J. Immunol. 28:1-11(1998).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM 1), AND VARIANTS LEU-44; THR-50; THR-52;
 RP ARG-54; ALA-57; GLY-75; ASP-95; LEU-96; ASN-100; ARG-107; GLY-109 AND
 RP VAL-131.
 RC TISSUE=Brain;
 RX MEDLINE=20053880; PubMed=10585853;
 RA Sano S.-I., Ohnishi H., Kubota M.;
 RT "Gene structure of mouse BIR/SHPs-1.";
 RL Biochem. J. 344:667-675(1999).
 RN [5]

RP SEQUENCE FROM N.A.
 RX MEDLINE=1638749; PubMed=11780052;
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
 RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
 RA Bailely O.P., Bird C.P., Blakey S.E., Bridgman M., Brown A.J.,
 RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
 RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
 RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
 RA Coulson A.G., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
 RA Graffam D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
 RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
 RA Levasalho M.H., Leverisa M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
 RA Marsh V.L., Martin S.L., McCormachie L.J., McKay K., McMurray A.A.,
 RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
 RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
 RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showgreen R., Sims S.,
 RA Snace C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
 RA Swann R.M., Syammore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
 RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
 RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,
 RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
 RA Rogers J.;
 RT "The DNA sequence and comparative analysis of human chromosome 20.";
 RL Nature 414:865-871(2001).
 [6]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND VARIANTS LEU-44; THR-50;
 RP THR-52; ARG-54; ALA-57; GLY-75; ASP-95; LEU-96; ASN-100; ARG-107;
 RP GLY-109 AND VAL-131.
 RC TISSUE=Brain, Kidney, and Skin;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Straube R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Kraemer R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heien F.,
 RA Diatchenko L., Marusha K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Uedin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raba S.S., Loughran N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.V., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whitting R.W., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butcherfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.W., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 [7]
 RP FUNCTION AND INTERACTIONS WITH FYB, SCAP2 AND PTK2B.
 RX MEDLINE=99401000; PubMed=10469599;
 RA Tjims J.F., Swanson K.D., Marie-Cardine A., Raab M., Rudd C.E.,
 RA Schraven B., Neel B.G.;
 RT "SHPS-1 is a scaffold for assembling distinct adhesion-regulated
 RT multi-protein complexes in macrophages.";
 RL Curr. Biol. 9:927-930(1999).
 [8]
 RP PHOSPHORYLATION BY JAK2, AND INTERACTIONS WITH PTPN11 AND JAK2.
 RX MEDLINE=00428747; PubMed=10842184;
 RA Scofega M.R., Argersinger L.S., Wang H., Ullrich A., Carter-Su C.;
 RT "Negative regulation of growth hormone receptor/JAK2 signaling by
 RT signal regulatory protein alpha.";
 RL J. Biol. Chem. 275:28222-28229(2000).
 [9]
 RP FUNCTION AND INTERACTION WITH CD47.
 RX MEDLINE=21400825; PubMed=11509594;

RA Latour S., Tanaka H., Demeure C., Mateo V., Rubio M., Brown E.J.,
 RA Maliszewski C., Lindberg F.P., Oldenburg A., Ullrich A.,
 RA Maliszewski G., Sarfati M.;
 RT "Bidirectional negative regulation of human T and dendritic cells by
 RT CD47 and its cognate receptor signal-regulator protein-alpha.";
 RT down-regulation of IL-12 responsiveness and inhibition of dendritic
 RT cell activation.";
 RL J. Immunol. 167:2547-2554(2001).
 CC -1- FUNCTION: Immunoglobulin-like cell surface receptor for CD47. Acts
 CC as docking protein and induces translocation of PTPN6, PTPN11 and
 CC other binding partners from the cytosol to the plasma membrane.
 CC Supports adhesion of cerebellar neurons, neurite outgrowth and
 CC glial cell attachment. May play a key role in intracellular
 CC signaling during synaptogenesis and in synaptic function (By
 CC similarity). Involved in the negative regulation of receptor
 CC tyrosine kinase-coupled cellular responses induced by cell
 CC adhesion, growth factors or insulin. Mediates negative regulation
 CC of phagocytosis, mast cell activation and dendritic cell
 CC activation. CD47 binding prevents maturation of immature dendritic
 CC cells and inhibits cytokine production by mature dendritic cells.
 CC -1- SUBUNIT: Binds PTPN11 when tyrosine-phosphorylated, except in
 CC macrophages, where it primarily binds PTPN6. Binds GPR2 in vitro.
 CC Binds FGR (By similarity). Binds JAK2 irrespective of its
 CC phosphorylation status and forms a stable complex. Binds SCAP1
 CC and/or SCAP2. The resulting complex recruits FYB. Binds PTK2B.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- ALTERNATIVE PRODUCTS: Type I membrane protein.
 CC Event=Alternative splicing; Named isoforms=3;
 CC Name=1;
 CC IsoId=P78324-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=P78324-2; Sequence=VSP_007030;
 CC Note=No experimental confirmation available;
 CC Name=3;
 CC IsoId=P78324-3; Sequence=VSP_007029;
 CC Note=No experimental confirmation available;
 CC TISSUE SPECIFICITY: Ubiquitous. Highly expressed in brain.
 CC Detected on myeloid cells, but not T cells. Detected at lower
 CC levels in heart, placenta, lung, testis, ovary, colon, liver,
 CC small intestine, prostate, spleen, kidney, skeletal muscle and
 CC pancreas.
 CC -1- PTM: N-glycosylated.
 CC -1- PTM: Phosphorylated on tyrosine residues in response to
 CC stimulation with EGF, growth hormone, insulin and PDGF.
 CC Dephosphorylated by PTPN11.
 CC -1- SIMILARITY: Contains 2 immunoglobulin-like C1-type domains.
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; D86043; BAA12974.1; -;
 DR EMBL; Y10375; CAA71403.1; -;
 DR EMBL; Y11047; CAA71944.1; ALT INIT.
 DR EMBL; AB023430; BAA87929.1; -;
 DR EMBL; AC004832; AAF19260.1; -;
 DR EMBL; AL034562; CAB38874.1; -;
 DR EMBL; AL049634; CAB46662.1; ALT_SEQ.
 DR EMBL; AL117335; CAC12723.1; -;
 DR EMBL; BC026692; AAH26692.1; -;
 DR EMBL; BC033092; AAH33092.1; -;
 DR EMBL; BC038510; AAH38510.1; -;
 DR PIR; JCS287; JCS287.
 DR Genew; HGNC; 9662; PTPN11.
 DR MIM; 602461; -;
 DR GO; GO:0005886; C:plasma membrane; TAS.
 Query Match 4.2%; Score 143.5; DB 1; Length 503;

Best Local Similarity 21.8%; Pred. No. 0.059;
Matches 87; Conservative 54; Mismatches 166; Indels 93; Gaps 17;

QY KEVSVKRVTPDKPKQWKKLPLHLTLPLAL-----QVAGSGL-TLLEKTKTKLQEV 312
DB ELVOIQPDSDSVSAAGSALHCTVSLIVGPIQWFRGAPRELTYNQKHFF--- 87
QY NLVVMRATQLOKNTCEVWGPTSPKMLSLLENKKAQVSKREKRVWLNPEAGMOCIL 372
DB -----PRVTVSSESTRKEMNDPSISN-----ITPADACTTYCVK 123
QY SDGQVLE-----SNIKVLPVSTVPYPCAPAPKSCDKTKTCELLGGSVLPFPKP 426
DB FRKSPDTEFKSGAGTSLVSAKPSAPV-----VSGPAARATP--- 161
QY KDTLMISTREVTGVVVDVSHEDPEVKMNVVDGEVNAATK--PRE-----QYNSTYR 480
DB -----QHTVSFTESHGSPRD--ITLWFKNGNELSDFTVNDPVGESVYSIHSTAK 213
QY VSVVTLVLDHOMLNGKEYKCKVSNKALPA-PIEKT--ISKAKGQPREPOVYTLPPSRDEL 537
DB VVLRREVDHSGVI-----CEVAHVTLQGDPLRGITANLSEITIRVPPLLEV-TQGPVPAE- 265
QY TRKQVSLTCLVKGFPSPDIQVWESNGQPNENYKTPPEVLDSGFFLYSKLTVDKSRWQ 597
DB --NQVNTVCQVRFKFPQRLQLTWLENGVNSRTETASTVTEKDGTYMMSVLLVNSAHR 323
QY 598 QGNVFSQSVHME---ALHNHTOKSLSLSPGLQDCECAE 634
DB 324 DVKLTQVEHGDGPAVSKSHDLK-VSAHPKEQGSNTAAE 362

RESULT 95
CAML_MOUSE
ID CAML_MOUSE STANDARD; PRT; 1260 AA.
AC P11627;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Neural cell adhesion molecule L1 precursor (N-CAM L1).
GN L1CAM OR CAML1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TextID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Brain;
RX MEDLINE=88318924; PubMed=3412448;
RA Moos M., Tacke R., Scherer H., Teplow D., Fruh K., Schachner M.;
RT "Neural adhesion molecule L1 as a member of the immunoglobulin
superfamily with binding domains similar to fibronectin.";
RL Nature 334:701-703(1988).
RT -1- FUNCTION: Cell adhesion molecule with an important role in the
development of the nervous system. Involved in neuron-neuron
adhesion, neurite fasciculation, outgrowth of neurites, etc. Binds
to axonin on neurons.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily.
CC -1- SIMILARITY: Contains 6 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 5 fibronectin type III domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.ebi.ac.uk/announcements>
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; X12875; CAA31368.1; -;
DR PIR; S05479; S05479; -;

DR HSP; P20241; ICFB.
DR MCD; MGI:96721; L1cam.
DR GO; GO:0007411; P:axon guidance; IMP.
DR InterPro; IPR006957; FN_III-like.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR007110; IG_III.
DR InterPro; IPR003598; IG_C2.
DR Pfam; PF00047; Ig_6.
DR Pfam; PF00060; FN3_2.
DR SMART; SM00408; IGc2_5.
DR PROSITE; PSS0835; IG_LIKE; 6.
DR Neurogenesis; Cell adhesion; Developmental protein; Glycoprotein;
KW Transmembrane; Repeat; Immunoglobulin domain; Signal.
FT SIGNAL 1
FT CHAIN 20 1260
FT DOMAIN 20 1123
FT TRANSMEM 1124 1146
FT DOMAIN 1147 1260
FT DOMAIN 135 130
FT DOMAIN 138 225
FT DOMAIN 239 327
FT DOMAIN 332 419
FT DOMAIN 424 506
FT DOMAIN 517 600
FT DOMAIN 827 896
FT DOMAIN 932 1094
FT DOMAIN 1032 1094
FT SITE 553 555
FT SITE 562 564
FT DISULFID 57 113
FT DISULFID 157 208
FT DISULFID 263 311
FT DISULFID 353 403
FT DISULFID 447 496
FT DISULFID 538 590
FT CARBOHYD 100 100
FT CARBOHYD 202 202
FT CARBOHYD 246 246
FT CARBOHYD 293 293
FT CARBOHYD 432 432
FT CARBOHYD 478 478
FT CARBOHYD 489 489
FT CARBOHYD 504 504
FT CARBOHYD 587 587
FT CARBOHYD 670 670
FT CARBOHYD 725 725
FT CARBOHYD 776 776
FT CARBOHYD 824 824
FT CARBOHYD 848 848
FT CARBOHYD 875 875
FT CARBOHYD 968 968
FT CARBOHYD 978 978
FT CARBOHYD 1022 1022
FT CARBOHYD 1030 1030
FT CARBOHYD 1073 1073
FT CARBOHYD 1107 1107
SQ SEQUENCE 1260 AA; 140968 MW; 22BE57001CB2A538 CRC64;

Query Match 4.2%; Score 143.5; DB 1; Length 1260;
Best Local Similarity 18.9%; Pred. No. 0.2;
Matches 122; Conservative 94; Mismatches 252; Indels 179; Gaps 29;

QY 8 RHLLV---LQALPA-ATQGNRVVLGKGDVLTCTASQKSIQFHWKNSQIKI 61
DB 408 OHGLLANAVIYVQPARILTRDQTYMAVESTAYLLCKARGAPVPSQVMDREGTV 467
QY 62 LQNGSFLTGPBKADRADRSRWLDQGNFPIIKLKIESDYTYICEVEDQKEVQL 121
DB 468 LQDERFF-----PYANGTISIDLDLQANDYGRVFCQADQNNVTLL 508
QY 122 VFGLTANSDTHLQ-----QSLTTLTLESPGSGSPVQCR-----SPRGKN-- 162

ID SRB1 HUMAN STANDARD; PRT; 398 AA.
 AC O00241; O8TB12; O9H105; O9Y4V0;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DE 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Signal-regulatory protein beta-1 precursor (SIRP-beta-1).
 GN SIRPB1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Placenta;
 RX MEDLINE=97215901; PubMed=9062191;
 RA Karltonenkov A., Chen Z., Sures I., Wang H., Schilling J.,
 RT Ullrich A.;
 RT "A family of proteins that inhibit signalling through tyrosine kinase
 RT receptors.";
 RL Nature 386:181-186(1997).
 RL [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RX MEDLINE=21638749; PubMed=11780052;
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
 RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
 RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
 RA Buck D., Buttrill W.D., Butler A.P., Carder C., Carter N.P.,
 RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
 RA Copley S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
 RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
 RA Grahm D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
 RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
 RA Lehesvallo M.H., Levenson M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
 RA Marsh V.L., Martin S.L., McConachie L.J., McInay K., McMurtry A.A.,
 RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
 RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsey H.,
 RA Rice C.M., Rose M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,
 RA Stuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulten J.E.,
 RA Swann R.M., Symmore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
 RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
 RA Whitehead S.L., Whitelaker P., Willey D.L., Williams L., Williams S.A.,
 RA Wilmink L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
 RT Rogers J.;
 RT "The DNA sequence and comparative analysis of human chromosome 20.";
 RL Nature 414:865-871(2001).
 RL [3]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Skin;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusik K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.P., Scheetz T.E.,
 RA Brownstein M.J., Ueffing T.B., Toehlyuk L.S., Carninci P., Prange C.J.,
 RA Raha S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bobak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richardson S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalek U., Smallie D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [4]
 RP INTERACTION WITH TYROBP AND SYK.
 RX PubMed=11169422.
 RA Tomasello E., Cant C., Buehring H.-J., Vely F., Ande P., Seifert M.,
 RA Ullrich A., Vlyer E.;
 RT "Association of signal-regulatory proteins beta with KAPAP/DAP-12.";
 RL Eur. J. Immunol. 30:2147-2156(2000).
 RN [5]
 RP INTERACTION WITH TYROBP, N-GLYCOSYLATION, AND TISSUE SPECIFICITY.
 RX MEDLINE=20072721; PubMed=10604985;
 RA Dietrich J., Cella M., Seifert M., Buehring H.-J., Colonna M.;
 RT "Signal-regulatory protein beta 1 is a DAP12-associated activating
 RT receptor expressed in myeloid cells.";
 RL J. Immunol. 164:9-12(2000).
 CC -1- FUNCTION: Immunoglobulin-like cell surface receptor involved in
 CC the negative regulation of receptor tyrosine kinase-coupled
 CC signaling processes. Participates also in the recruitment of
 CC tyrosine kinase SYK.
 CC -1- SUBUNIT: Interacts with TYROBP. This interaction results in the
 CC recruitment of SYK.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=O00241-1; Sequence=displayed;
 CC Name=2;
 CC IsoId=O00241-2; Sequence=VSP_007026;
 CC Note=No experimental confirmation available;
 CC -1- TISSUE SPECIFICITY: Detected in monocytes and dendritic cells.
 CC -1- PTM: N-glycosylated.
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Y10376; CAA71404.1; -;
 DR EMBL: AL049634; CAB4661.2; -;
 DR EMBL: AL138804; CAC17540.1; -;
 DR EMBL: BC025286; AAH25286.1; -;
 DR Genew; HGNC:15928; SIRPB1.
 DR MIM; 603889; -;
 DR GO; GO:0005887; C:integral to plasma membrane; TAS.
 DR GO; GO:0007166; P:cell surface receptor linked signal transdu. . .; TAS.
 DR InterPro; IPR007110; Ig_1like.
 DR InterPro; IPR003597; Ig_c1.
 DR Pfam; PF00047; Ig; 3.
 DR SMART; SM00407; IgC1; 2.
 DR PROSITE; PSS0835; IG_LIKE; 3.
 KW Repeat; Signal; Transmembrane; Immunoglobulin domain; Glycoprotein;
 KW Alternative splicing.
 FT FT SIGNAL 1 26
 FT CHAIN 27 398
 FT DOMAIN 27 371
 FT DOMAIN 372 392
 FT TRANSMEM 393 398
 FT DOMAIN 27 136
 FT DOMAIN 147 246
 FT DOMAIN 253 347
 FT DISULFID 54 120
 FT DISULFID 169 227
 FT DISULFID 272 320
 FT CARBOHYD 244 244
 FT CARBOHYD 269 269
 FT CARBOHYD 291 291
 FT VARSPPLIC 145 361
 /FTId=VSP_007026.
 POTENTIAL.
 SIGNAL-REGULATORY PROTEIN BETA-1.
 EXTRACELLULAR (POTENTIAL).
 CYTOPLASMIC (POTENTIAL).
 IG-LIKE V-TYPE.
 IG-LIKE C1-TYPE 1.
 IG-LIKE C1-TYPE 2.
 POTENTIAL.
 POTENTIAL.
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 Missing (in isoform 2).
 /FTId=VSP_007026.


```

FT DOMAIN 234 326 IG-LIKE C2-TYPE 3.
FT DOMAIN 331 416 IG-LIKE C2-TYPE 4.
FT DOMAIN 426 522 FIBRONECTIN TYPE-III 1.
FT DOMAIN 525 618 FIBRONECTIN TYPE-III 2.
FT DOMAIN 619 716 FIBRONECTIN TYPE-III 3.
FT DOMAIN 722 816 FIBRONECTIN TYPE-III 4.
FT DOMAIN 840 940 FIBRONECTIN TYPE-III 5.
FT DOMAIN 941 1042 FIBRONECTIN TYPE-III 6.
FT DISULFID 61 117 BY SIMILARITY.
FT DISULFID 161 212 BY SIMILARITY.
FT DISULFID 261 310 BY SIMILARITY.
FT DISULFID 352 400 BY SIMILARITY.
FT CARBOHYD 94 94 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 318 318 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 478 478 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 628 628 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 702 702 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 168 168 M -> T (in oesophageal carcinoma).
FT VARIANT 201 201 R -> G.
FT VARIANT 201 201 /FTID=VAR_003910.
FT VARIANT 1375 1375 P -> H (in a colorectal carcinoma).
FT CONFLICT 138 138 /FTID=VAR_003911.
FT CONFLICT 233 329 MISSING (IN REF. 3).
FT CONFLICT 421 421 MISSING (IN REF. 3).
FT CONFLICT 421 421 MISSING (IN REF. 3).
FT CONFLICT 1447 AA; 158456 MM; 4A8612765ED0471F CRC64;
SEQUENCE

```

```

Query Match 4.1%; Score 141.5; DB 1; Length 1447;
Best Local Similarity 18.7%; Pred. No. 0.32;
Matches 131; Conservative 82; Mismatches 225; Indels 263; Gaps 33;

```

```

QY 34 GDTYELCTTSQKSIQFHMKNQKILKQSGFLTKGPKSLKDRADRSRLMDQGNP 93
DB 154 GDTYELCTTSQKSIQFHMKNQKILKQSGFLTKGPKSLKDRADRSRLMDQGNP 93
QY 94 LIINKLIEDSDTYCEVED-----QKEVOLL-----VFGLTANSDFHLQSGS 138
DB 197 LQISRLPGDGLGRSARNAASRTGNEAVRLSDPLGRQLYFQRPNNVVALISGKO 256
QY 139 LTL---TLESPPGS-----SPSVQCRSPRGKNIQSGKTLVSQLELDGSGTWTCTYLO 189
DB 257 AVLECCSGYPRPFMTLRGEVYQLRSKK-YSLGGSNLISVNTDDSGMTYCVTVYK 315
QY 190 QKKVEFIDIVLA--FQKASSIVYKKEGQVEFSPPLA----- 226
DB 316 NENISAGAEILTVLPWFLNHPNLVAVESMDIEFFECTSGKPVPTVMMKGDVVI 375
QY 227 -----FTVEKLTGSGELMNG--AERASSKSMITFDLKNKEVSKVYTOPKLOMG 275
DB 376 YFOVIGGSNLRITLVKXSDGEFYOCVAENENG-----NAQTSAQILVPRPAIPSS 425
QY 276 KKLPLHLTLPOAL-----PQYAGSGNT--TLAEATKGKLGHEVNLVVM 317
DB 426 SVLP---SARDVPLVVSRRFVRLSMRPPAEAKGNIQTFFVFSRGDRERLNTOP 482
QY 318 RAFOQL-KNLTCEV-----WGP-----TSPKMLSLKLENKEAKVSKRE 355
DB 483 GSIQLTGNLKPFAEMYTFRVVAVNEMGSGESSQPIKATQBELQVPPREVLQA----- 536
QY 356 KPVVNLNPEAGMOCCLSDSGQVLLESNIKLPFTWSTPVPPEAPPE---KSCDKTHRCPE 412
DB 537 -----VSTSPSTLITLWEPAPYANGVQVRLPCTEVSIGKE 573
QY 413 -----LIG-----GPSVFLFPKPKXTLMSIRTPETCVV- 443
DB 574 QNIEVDGLSYKLBJKKFTESLAFLLAVNRKGPV-----STDDITVVTLS 619
QY 444 DVSHEDPE-----VKFNW-----YVDGEVHNAKTRPRE----QYNST 478
DB 620 DVPSAPQONVSLLEVNSRSIKVSWLPPPSGTQNGFITGYKIRHRTTRRGEMETLEPNL 679

```

```

QY 479 YRVSVTLVTHQDMLNGKCKVSNKALPAPIEKTISKAGOPREOVYTLPPSRDEL 538
DB 680 WYLFTE-----KSGQSPQVSAMTV-----NQTGPP--SNWTAETPEMDLD 721
QY 539 KNOVSLTCLVKGFPSPDIKAVESNQPENN---YKTPPV 576
DB 722 ESQVP-----DQPSGLHV-----RPTNCIIMSWTPEL 749

RESULT 99
SNE2 HUMAN
ID SNE2 HUMAN STANDARD; PRT; 6885 AA.
AC Q8WXH0; Q8N1S3; Q8NF49; Q8TER7; Q8MWW3; Q8MWW4; Q8MWW5; Q8WXH1;
AC Q8NUS0; Q9UP04; Q9Y2L4; Q9Y4L1;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Nesprin 2 (Nuclear envelope spectrin repeat protein 2) (Syne-2)
DE (Synaptic nuclear envelope protein 2) (Nucleus and actin connecting
DE element protein) (NUANCE protein).
GN SYNE2 OR NUA OR KIA1011.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A. (ISOFORMS 1 AND 9), FUNCTION, CHARACTERIZATION, AND
INTERACTION WITH F-ACTIN.
RX MEDLINE=22113122; PubMed=12118075;
RA Zhen Y.-Y., Libotte T., Munck M., Noegel A.A., Korenbaum E.;
RT "NUANCE, a giant protein connecting the nucleus and actin
RT cytoskeleton."
RL J. Cell Sci. 115:3207-3222(2002).
[2]
SEQUENCE FROM N.A. (ISOFORMS 4; 5 AND 7).
RX MEDLINE=21652858; PubMed=11792814;
RA Zhang Q., Skepper J.N., Yang F., Davies J.D., Hegyi L., Roberts R.G.,
RA Weisberg P.L., Ellis J.A., Shanahan C.M.;
RT "Nesprins: a novel family of spectrin-repeat-containing proteins that
RT localize to the nuclear membrane in multiple tissues."
RL J. Cell Sci. 114:4485-4498(2001).
[3]
SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=22296983; PubMed=12408964;
RA Zhang Q., Regnault C., Greener M.J., Shanahan C.M., Roberts R.G.;
RT "The nesprins are giant actin-binding proteins, orthologous to
RT Drosophila melanogaster muscle protein MSP-300."
RL Genomics 80:473-481(2002).
[4]
SEQUENCE FROM N.A. (ISOFORM 6).
RX MEDLINE=12508121;
RA Hellig R., Eckenberg R., Petit J.-L., Fonknechten N., Da Silva C.,
RA Cattoletto L., Levy M., Barbe V., De Bernardinis V., Ureta-Vidal A.,
RA Pelleterio E., Vico V., Anthonard V., Rowen L., Madan A., Qin S.,
RA Sun H., Du H., Pepin K., Artigianave F., Robert C., Cruaud C.,
RA Bruehl T., Jallion O., Friedlander L., Samson G., Brotier P.,
RA Cure S., Segurens B., Aniere F., Samain S., Crespeau H., Abbasi N.,
RA Alich N., Boscus D., Dickhoff R., Dora W., Dubois I., Friedman C.,
RA Gouvenoux M., James R., Madan A., Mailey-Estrada B., Mangenot C.,
RA Martins N., Menard M., Ozlas S., Ratcliffe A., Shaffer T., Trask B.,
RA Vacherie B., Bellemere C., Belser C., Bernard-Gonet M.,
RA Butoise-Mavel D., Bourard M., Brizet-Silla S., Combette S.,
RA Dufosse-Laurent V., Perron C., Lechaplaix C., Louessee C., Museliet D.,
RA Magdeleat G., Pateau B., Petit E., Strvain-Itukiewicz P., Trybou A.,
RA Vega-Czarny N., Bataille E., Bluet E., Bordaie I., Dubois M.,
RA Dumont C., Guerin T., Haffray S., Hammedi R., Muanga J., Pellouin V.,
RA Robert D., Wunderle E., Gauguet G., Roy A., Sainte-Marthe L.,
RA Verdier J., Verdier-Discala C., Hillier L.W., Fulton L., McPherson J.,

```

RA Macsuda F., Wilson R., Scarpelli C., Gyapay G., Wincker P., Saurin W.,
RA Queller F., Waterston R., Hood L., Weissensbach J.;
RT "The DNA sequence and analysis of human chromosome 14.";
RL Nature 421:601-607(2003).
RN [6]
RP SEQUENCE FROM N.A. (ISOFORM 8).
RC TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Collins E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shmolen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA DiCicco L., Marusik K., Farmer A., Rubin G.M., Hong L.,
RA Stepien M., Soares M.B., Bonaldo M.F., Caesavert T.L., Schertz T.E.,
RA Brownstein M.J., Uediri T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rana S.S., Loguanello N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bonak S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettner M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting R.W., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.T., Skalska U., Smalios D.E.,
RA Schermer A., Schein J.E., Jones S.J.M., Marra W.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RL [7]
RP SEQUENCE FROM N.A. (ISOFORM 3), AND SEQUENCE OF 1-956 AND 5133-6885
RP FROM N.A.
RC TISSUE=Spleen, and Tongue;
RA Kiyuna H., Takano J., Nomura N., Kikuno R., Nagase T., Ohara O.,
RA Nitomiyu K., Maegatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
RA Furuya T., Takahashi M., Kikawa E., Omura Y., Abe K., Kamihara K.,
RA Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie Y.,
RA Otsuki T., Sato H., Wakamatsu A., Ishi S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Masuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,
RA Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA Suzuki Y., Sugano S., Nagahata K., Maeno Y., Nagai K., Isogai T.,
RT "NEDO human cDNA sequencing project.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
RN [8]
RP SEQUENCE OF 5709-6885 FROM N.A. (ISOFORM 2).
RN TISSUE=Brain;
RC MEDLINE=99246063; PubMed=10231032;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirosewa M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIII
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:63-70(1999).
RN [9]
RP REVISIONS.
RC TISSUE=Brain;
RX MEDLINE=22158633; PubMed=12168954;
RA Nakajima D., Okazaki N., Yamakawa H., Kikuno R., Ohara O., Nagase T.;
RT "Construction of expression-ready cDNA clones for KIAA genes: manual
RT curation of 330 KIAA cDNA clones.";
RL DNA Res. 9:99-106(2002).
RN [10]
RP SEQUENCE OF 5754-6885 FROM N.A.
RX MEDLINE=21546917; PubMed=11230166;
RA Wiemann S., Weil B., Wellenreuther R., Gaessenhuber J., Glasel S.,
RA Ansoyge W., Boecher M., Bloecher H., Bauewachs S., Blum H.,
RA Lander J., Duesethoelt A., Beyer A., Koehler K., Strick N.,
RA Meves H.-W., Ottensmelder B., Oelmaier B., Tampe J., Heubner D.,
RA Wandurt R., Korn B., Klein W., Poustka A.;
RT "Towards a catalog of human genes and proteins: sequencing and
RT analysis of 500 novel complete protein coding human cDNAs.";
RL Genome Res. 11:422-435(2001).
CC -1- FUNCTION: Involved in the maintenance of nuclear organization and

```

CC structural integrity. Probable anchoring protein which tethers the
CC nucleus to the cytoskeleton. Connects nuclei to the cytoskeleton
CC by interacting with the nuclear envelope and with F-actin in the
CC cytoplasm.
CC
CC -1- SUBUNIT: Interacts with F-actin via its N-terminal domain.
CC
CC -1- SUBCELLULAR LOCATION: Type IV membrane protein (Potential). The
CC largest part of the protein is cytoplasmic, while its C-terminal
CC part is associated with the nuclear envelope, most probably the
CC outer nuclear membrane. Remains associated with the nuclear
CC envelope during its breakdown in mitotic cells.
CC
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=9;
CC
CC Name=1;
CC IsoId=Q8WXH0-1; Sequence=Displayed;
CC
CC Name=2;
CC IsoId=Q8WXH0-2; Sequence=VSP_007164, VSP_007166;
CC Note=No experimental confirmation available;
CC
CC Name=3;
CC IsoId=Q8WXH0-3; Sequence=VSP_007155;
CC Note=Produced by exon skipping that results in a frameshift. No
CC experimental confirmation available;
CC
CC Name=4; Synonyms=Beta;
CC IsoId=Q8WXH0-4; Sequence=VSP_007156;
CC
CC Name=5; Synonyms=Alpha;
CC IsoId=Q8WXH0-5; Sequence=VSP_007157, VSP_007164, VSP_007165;
CC
CC Name=6;
CC IsoId=Q8WXH0-6; Sequence=VSP_007158, VSP_007165, VSP_007166;
CC Note=No experimental confirmation available;
CC
CC Name=7; Synonyms=Gamma;
CC IsoId=Q8WXH0-7; Sequence=VSP_007154, VSP_007163;
CC
CC Name=8;
CC IsoId=Q8WXH0-8; Sequence=VSP_007161, VSP_007162;
CC Note=No experimental confirmation available;
CC
CC Name=9; Synonyms=NUANCE-N-33;
CC IsoId=Q8WXH0-9; Sequence=VSP_007159, VSP_007160;
CC
CC -1- TISSUE SPECIFICITY: Widely expressed, with higher level in kidney,
CC adult and fetal liver, stomach and placenta. Weakly expressed in
CC skeletal muscle and brain. Isoform 5 is highly expressed in
CC pancreas, skeletal muscle and heart.
CC
CC -1- DOMAIN: The Klarsicht domain mediates the nuclear envelope
CC targeting.
CC
CC -1- SIMILARITY: Belongs to the neopterin family.
CC
CC -1- SIMILARITY: Contains 1 actin-binding domain.
CC
CC -1- SIMILARITY: Contains 2 calponin-homology (CH) domains.
CC
CC -1- SIMILARITY: Contains 1 Klarsicht domain.
CC
CC -1- SIMILARITY: Contains 1 leucine-rich (LRR) repeats.
CC
CC -1- SIMILARITY: Contains 9 spectrin repeats.
CC
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial/
CC entities requires a license agreement (See http://www.isb-ebc.ch/announce/
CC or send an email to license@isb-ebc.ch).
CC
CC
CC EMBL; AF435010; AAL33547.1; -
CC EMBL; AF435011; AAL33548.1; -
CC EMBL; AY061757; AAL33800.1; -
CC EMBL; AY061758; AAL33801.1; -
CC EMBL; AY061759; AAL33802.1; -
CC EMBL; AF495911; AAN60443.1; -
CC EMBL; AL117404; CAB55905.1; -
CC EMBL; AL162832; -; NOT_ANNOTATED_CDS.
CC EMBL; ALJ35094; -; NOT_ANNOTATED_CDS.
CC
CC
CC Query Match 4.1%; Score 141; DB 1; Length 6885;
CC Best Local Similarity 25.6%; Pred. No. 2.6;
CC Matches 83; Conservative 59; Mismatches 128; Indels 54; Gaps 15;

```


THIS PAGE BLANK (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 3, 2004, 13:00:24 ; Search time 14.8924 Seconds
(without alignments)
4185.504 Million cell updates/sec

Title: SE07
Perfect score: 3414
Sequence: 1 MNRGVPRHLLVLQLALLP.....DETCAEADGELDGLMTTDP 648

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 125 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2015	59.0	458	1 RWHUT4	T-cell surface gly
2	1867	54.7	432	1 RWCZT4	T-cell surface gly
3	1729	50.6	432	1 RWMQT4	T-cell surface gly
4	1211.5	35.5	330	1 GHU	Ig gamma-1 chain C
5	1205.5	35.3	374	2 S69339	Ig heavy chain V r
6	1203.5	35.3	255	4 S1866	Ig gamma-1 chain C
7	1171.5	34.3	234	2 PTO207	Ig gamma-3 chain C
8	1163	34.1	377	2 A23511	Ig gamma-3 chain C
9	1161	34.0	377	2 A60764	Ig gamma-3 chain C
10	1150.5	33.7	459	2 A46254	CD4 precursor - ra
11	1143	33.5	289	1 G3HUW1	Ig gamma-3 heavy c
12	1115.5	32.7	326	1 G2HU	Ig gamma-2 chain C
13	1105.5	32.4	327	1 G4HU	Ig gamma-4 chain C
14	1099	32.2	432	2 S30193	T-cell surface gly
15	1000.5	29.3	457	2 A27449	T-cell surface gly
16	993.5	29.1	457	1 RWMST4	T-cell surface gly
17	938.5	27.5	398	1 G3MSM	Ig gamma-3 chain C
18	928.5	27.2	393	1 G1MSM	Ig gamma-1 chain C
19	910	26.7	323	1 GHRB	Ig gamma-1 chain C
20	902.5	26.4	339	1 G2MSAM	Ig gamma-2a chain
21	901.5	26.4	328	2 I47160	Ig gamma-2b chain
22	901.5	26.4	328	2 I47159	Ig gamma-2a chain
23	894	26.2	277	2 I47162	Ig gamma-2 chain C
24	879.5	25.8	329	1 G2GP	Ig gamma-2 chain C
25	879.5	25.8	405	1 G2MSBM	Ig gamma-2b chain
26	876.5	25.7	328	2 I47161	Ig gamma-1 chain C
27	876.5	25.6	328	2 I47158	Ig gamma-1 chain C
28	870.5	25.5	470	2 S22080	Ig heavy chain pre
29	843	24.7	472	2 S31459	Ig gamma-1 chain -

30	840.5	24.6	329	1 G3MSC	Ig gamma-3 chain C
31	836	24.5	308	2 C30554	Ig heavy chain C r
32	830.5	24.3	444	2 PC4436	monoclonal antibod
33	829	24.3	326	2 PS0017	Ig gamma-1 chain C
34	826.5	24.2	333	2 PS0018	Ig gamma-2b chain
35	823.5	24.1	469	2 S37483	Ig gamma-2a chain
36	822.5	23.9	329	2 S00847	Ig gamma-2c chain
37	814.5	23.6	324	1 G1MS	Ig gamma-1 chain C
38	807	23.5	474	1 G2MS11	Ig gamma-2b chain
39	801.5	23.5	330	1 G2MSA	Ig gamma-2a chain
40	798	23.4	446	2 S40295	Ig gamma-2a chain
41	792	23.2	322	2 PS0019	Ig gamma-2a chain
42	788	23.1	335	1 G2MSAB	Ig gamma-2a chain
43	779.5	22.8	475	2 S01321	Ig gamma-2b chain
44	768	22.5	327	2 S06611	Ig gamma-2 chain C
45	702	20.6	180	2 I46732	Ig gamma heavy cha
46	574.5	16.8	218	2 A36040	Ig heavy chain V-I
47	572.5	16.8	249	2 S69340	Ig heavy chain VHI
48	566	16.6	152	2 S14236	Ig gamma-1 chain C
49	401	11.7	549	2 S04845	Ig heavy chain pre
50	399.5	11.7	572	2 B46529	Ig y heavy chain (
51	389.5	11.4	627	2 S14683	Ig mu chain precu
52	374	11.0	548	2 S38864	Ig epsilon chain C
53	367.5	10.8	429	1 EHRT	Ig epsilon chain C
54	363.5	10.6	592	2 S25705	Ig epsilon chain C
55	363.5	10.6	592	2 S03186	Ig mu chain - sh
56	360.5	10.6	448	2 S00380	Ig gamma chain (cl
57	360	10.5	504	2 S00380	Ig gamma chain (cl
58	359.5	10.5	476	1 MHMSM	Ig mu chain C regi
59	359	10.5	474	2 S15590	Ig heavy chain - h
60	357.5	10.5	453	2 S37768	Ig mu chain C regi
61	357	10.5	71	2 I60082	CD4 receptor - hum
62	356	10.4	428	1 EHHU	Ig epsilon chain C
63	355.5	10.4	479	1 MHREM	Ig mu chain C regi
64	353.5	10.4	426	2 I36948	Ig epsilon chain -
65	352.5	10.3	391	1 MHHDUT	Ig mu heavy chain
66	351.5	10.3	473	1 MHHDUT	Ig mu chain C regi
67	348	10.2	455	1 MHMS	Ig mu chain C regi
68	348	10.2	455	2 A24976	Ig mu chain C regi
69	347	10.2	452	1 MHU	Ig mu chain C regi
70	346.5	10.1	423	1 EHMS	Ig epsilon chain C
71	346.5	10.1	454	1 MHY	Ig mu chain C regi
72	346	10.1	457	2 S03961	Ig mu chain C regi
73	346	10.1	458	1 MHRB	Ig heavy chain - n
74	345.5	10.1	577	2 I50731	Ig mu chain C regi
75	341.5	10.0	343	2 S25644	Ig mu chain C regi
76	340	10.0	450	1 MHMG	Ig mu chain C regi
77	331.5	9.7	112	2 B30503	Ig gamma-2a chain
78	326.5	9.6	684	2 S60266	novel antigen rece
79	324.5	9.5	453	2 C31933	Ig mu chain C regi
80	320.5	9.4	433	2 S31436	Ig epsilon chain -
81	306	9.0	299	1 AHRB	Ig alpha chain C r
82	305.5	8.9	99	2 S21461	T-cell surface gly
83	303	8.9	338	2 S09276	Ig alpha chain C r
84	303	8.9	461	1 HVRKC0	Ig mu chain C regi
85	301.5	8.8	438	1 HVRKC	Ig mu chain C regi
86	301.5	8.8	585	2 A46507	Ig alpha chain - c
87	298.5	8.7	342	2 S09270	Ig alpha chain C r
88	298	8.7	342	2 I47175	Ig alpha chain C r
89	297.5	8.7	438	1 HVRK2	Ig mu chain C regi
90	297	8.7	357	2 S09269	Ig alpha chain C r
91	295.5	8.7	347	2 S09274	Ig alpha chain C r
92	293.5	8.6	340	2 B22360	Ig alpha chain C r
93	293	8.6	367	1 MHCH	Ig mu chain C regi
94	289.5	8.5	393	1 HVRK26	Ig mu chain C regi
95	285.5	8.4	352	2 S09265	Ig alpha chain C r
96	285.5	8.4	357	2 S09265	Ig alpha chain C r
97	285	8.3	343	2 S09272	Ig alpha chain C r
98	284.5	8.3	340	2 I56230	Ig alpha-2 chain -
99	284	8.3	244	2 S12328	Ig heavy chain C r
100	283.5	8.3	352	2 S05500	Ig alpha-1 chain C
101	283.5	8.3	353	1 A1HU	Ig alpha-1 chain C
102	280.5	8.2	99	2 S21462	T-cell surface gly

103 278.5 8.2 474 2 I50830 Ig mu chain - Lep1
104 278 8.1 558 2 A45804 Ig mu chain C regi
105 277 8.1 339 2 S09264 Ig alpha chain C r
106 277 8.1 568 2 A34891 Ig heavy chain pre
107 275.5 8.1 88 2 A30503 Ig gamma-2b chain
108 274 8.0 360 2 S09271 Ig alpha chain C r
109 273.5 8.0 348 2 S09273 Ig alpha chain C r
110 271 7.9 340 1 A2HU Ig alpha-2 chain C
111 270 7.9 370 1 HVKCS Ig mu chain C regi
112 270 7.9 573 2 S12838 Ig mu chain precu
113 269.5 7.9 358 2 S09268 Ig alpha chain C r
114 266.5 7.8 357 2 S09267 Ig alpha chain C r
115 265 7.8 454 2 A4532 Ig mu chain C regi
116 258.5 7.6 342 2 A45966 Ig alpha chain C r
117 257 7.5 111 2 S43148 Ig upslon chain -
118 251.5 7.4 344 1 AHMS Ig alpha chain C r
119 247 7.2 1005 2 T1857 Ig heavy chain - c
120 243.5 7.1 220 2 C22360 Ig alpha-2 chain C
121 241.5 7.1 580 2 A46538 Ig heavy chain, se
122 239.5 7.0 335 2 S09275 Ig alpha chain C r
123 211 6.2 448 2 A46533 Ig heavy chain C r
124 209 6.1 115 2 I68731 Igb chain C4 regio
125 206 6.0 402 2 S20002 Ig heavy chain, se

ALIGNMENTS

RESULT 1
RMHTT4
T-cell surface glycoprotein CD4 precursor [validated] - human
N:Alternate names: T-cell surface antigen T4/Leu 3
C:Species: Homo sapiens (man)
C>Date: 28-May-1986 #sequence, revision 31-Dec-1988 #text, change 20-Apr-2001
C:Accession: A90872; A32722; A34194; A53287; I54176; I54297; A02109; A30039
R:Maddon, P.J.; Littman, D.R.; Godfrey, M.; Maddon, D.E.; Chess, L.; Axel, R.
Cell 42, 93-104, 1985
A:Title: The isolation and nucleotide sequence of a cDNA encoding the T cell surface pro
A:Reference number: A90872; MUID:8524948; PMID:2990730
A:Accession: A90872
A:Molecule type: mRNA
A:Residues: 1-25, 'N', 27-458 <MAD>
A:Experimental source: clone pT4B
R:Littman, D.R.; Maddon, P.J.; Axel, R.
Cell 55, 541, 1988
A:Title: Corrected CD4 sequence.
A:Reference number: A90907; MUID:89028665; PMID:3263213
A:Contents: annotation; revision to residue 26
R:Cammerlin, D.; Seed, B.
Cell 60, 747-754, 1990
A:Title: A CD4 domain important for HIV-mediated syncytium formation lies outside the vi
A:Reference number: A32722; MUID:90182664; PMID:2107024
A:Accession: A32722
A:Status: nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 26-426, 428-458 <CAM>
R:Garr, S.A.; Hemling, M.E.; Folea-Wasserman, G.; Sweet, R.W.; Anumula, K.; Barr, J.R.;
J. Biol. Chem. 264, 21286-21295, 1989
A:Title: Protein and carbohydrate structural analysis of a recombinant soluble CD4 recep
A:Reference number: A34194; MUID:90078232; PMID:2592374
A:Contents: disulfide bonds; carbohydrate-binding sites
A:Accession: A34194
A:Molecule type: protein
A:Residues: 26-394 <CAR>
R:Liedeman, S.; Demartino, J.A.; Daugherty, B.L.; Foeldvari, I.; Yellin, M.J.; Cleary, A
Mol. Immunol. 28, 1171-1181, 1991
A:Title: A single amino acid substitution in a common African allele of the CD4 molecule
A:Reference number: A53287; MUID:92072595; PMID:1961196
A:Accession: A53287
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 250-264, 'W', 266-280 <LED>
A:Note: sequence extracted from NCBI backbone (NCBIP:68249)

R:Edwards, M.C.; Gibbs, R.A.
Genomics 14, 590-597, 1992
A:Title: A human dimeric protein resulting from loss of an Alu.
A:Reference number: I54176; MUID:93052387; PMID:1330888
A:Accession: I54176
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-72 <RES>
A:Cross-references: GB:U47924; GB:M6525; GB:U72506; NID:g1633547; PIDN:AA16069.1; PID
R:Hodge, T.W.; Sasse, D.R.; McDougal, J.S.
Hum. Immunol. 30, 99-104, 1991
A:Title: Humans with OX74-epitope deficiency have a single nucleotide base change in th
A:Reference number: I54297; MUID:91216786; PMID:1708753
A:Accession: I54297
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-264, 'W', 266-458 <RE2>
A:Cross-references: GB:M35160; NID:g179143; PIDN:AA16069.1; PID:g179144
C:Comment: Macrophage tropic strains of HIV-1 bind to a complex of chemokine (C-C) recep
C:Genetics:
A:Gene: GDB:CD4
A:Cross-references: GDB:119767; OMIM:186940
A:Map position: 12pter-12p12
A:Intons: 16/3
C:Superfamily: T-cell surface glycoprotein CD4; immunoglobulin homology
C:Keywords: AIDS; duplication; glycoprotein; T-cell; transmembrane protein
F:1-25/Domain: signal sequence #status predicted <Sig>
F:26-458/Product: T-cell surface glycoprotein CD4 #status experimental <MAT>
F:34-111/Domain: immunoglobulin homology <IM1>
F:136-186/Domain: immunoglobulin homology #status atypical <IM2>
F:216-299/Domain: immunoglobulin homology <IM3>
F:321-372/Domain: immunoglobulin homology <IM4>
F:397-420/Domain: transmembrane #status predicted <TM>
F:421-458/Domain: intracellular #status predicted <INT>
F:41-109, 155-184, 328-370/Disulfide bonds: #status experimental
F:296, 325/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 59.0%; Score 2015; DB 1; Length 458;
Best Local Similarity 99.5%; Pred. No. 3, 1e-111;
Matches 392; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MNRGVFRRLLVLTQLALPAATQGNKVLGKKGDVETLTASQKKSIOFHKNSNOIK 60
Db 1 MNRGVFRRLLVLTQLALPAATQGNKVLGKKGDVETLTASQKKSIOFHKNSNOIK 60
QY 61 ILGNQSPFLTKGSPSKINDRADSRSLMDQGNFPLIKNLKIEDSDTYICEVEDQKEVOL 120
Db 61 ILGNQSPFLTKGSPSKINDRADSRSLMDQGNFPLIKNLKIEDSDTYICEVEDQKEVOL 120
QY 121 LVFGLTANSDDTHLLOQSILTLTLESPGSSPSVQGRSPRGKNTQGGKTLVSQLELDSDG 180
Db 121 LVFGLTANSDDTHLLOQSILTLTLESPGSSPSVQGRSPRGKNTQGGKTLVSQLELDSDG 180
QY 181 TWTCVTLQNKQKVEFKIDIVLAFOKASSIVYKKEGEQVEFSPPLAFTVEKLTGSGELMW 240
Db 181 TWTCVTLQNKQKVEFKIDIVLAFOKASSIVYKKEGEQVEFSPPLAFTVEKLTGSGELMW 240
QY 241 QMRASSSSKSWITFDLKNKEVSKYRTQPKLQMKKLPULHTLPALPOYAGSSGULTLA 300
Db 241 QMRASSSSKSWITFDLKNKEVSKYRTQPKLQMKKLPULHTLPALPOYAGSSGULTLA 300
QY 301 LEAKTGKLEHGVNLVYMRATOLQKNTLCEVWGPTSPKMLSLKLEKKAIVSKREKPVWV 360
Db 301 LEAKTGKLEHGVNLVYMRATOLQKNTLCEVWGPTSPKMLSLKLEKKAIVSKREKPVWV 360
QY 361 LNPBAGMOCILSDSGVLLSNIKVLPMTSTPV 394
Db 361 LNPBAGMOCILSDSGVLLSNIKVLPMTSTPV 394

RESULT 2
RMCTT4
T-cell surface glycoprotein CD4 - chimpanzee

N:Alternate names: T-cell surface antigen T4/Leu 3
C:Species: Pan troglodytes (chimpanzee)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Jul-1999
R:Accession: B32722; A46534
R:Camertini, D.; Seed, B.
Cell 60, 747-754, 1990
A:Title: A CD4 domain important for HIV-mediated syncytium formation lies outside the v
A:Reference number: A32722; MUID:90182664; PMID:2107024
A:Accession: B32722
A:Molecule type: mRNA
A:Residues: 1-432 <CM>
A:Cross-references: GB:M31135
R:Fomsgard, A.; Hirsch, V.M.; Johnson, P.R.
Eur. J. Immunol. 22, 2973-2981, 1992
A:Title: Cloning and sequences of primate CD4 molecules: diversity of the cellular recep
A:Reference number: A46534; MUID:93049640; PMID:11425921
A:Accession: A46534
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 3-399 <FOM>
A>Note: Sequence extracted from NCBI backbone (NCBIP:118332)
C:Comment: This protein is expressed on most thymocytes, on a subset of mature T-cells
C:Superfamily: T-cell surface glycoprotein CD4; immunoglobulin homology
F:1-432/Product: T-cell surface glycoprotein CD4 #status predicted <MAT>
F:1-371/Domains: extracellular #status predicted <EXT>
F:9-86/Domains: immunoglobulin homology <IM1>
F:111-161/Domains: immunoglobulin homology #status atypical <IM2>
F:191-274/Domains: immunoglobulin homology <IM3>
F:296-347/Domains: immunoglobulin homology <IM4>
F:372-395/Domains: transmembrane #status predicted <TM>
F:396-432/Domains: intracellular #status predicted <INT>
F:16-84,130-159,303-345/Disulfide bonds: #status predicted
F:271,300/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 54.7%; Score 1867; DB 1; Length 432;
Best Local Similarity 98.6%; Pred. No. 1.4e-102;
Matches 363; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 27 KVLGKKGDVLTCTASQKKSIOFHWKNSNQIKILGNQGSFLTKGSKLNDRAISRSL 86
DB 2 KVLGKKGDVLTCTASQKKSIOFHWKNSNQIKILGNQGSFLTKGSKLNDRAISRSL 61
QY 87 WDQGNFPLITIKNLKIDSDTYICEVEDQKEVOLLVGLTANSDTHLLOQSILTLTLESP 146
DB 62 WDQGNFPLITIKNLKIDSDTYICEVEDQKEVOLLVGLTANSDTHLLOQSILTLTLESP 121
QY 147 PGSSPSVQCRSPRGKNIQGGKTLVSQLELDSDGTMCTVLOKQKVEFKIDIVLAFOK 206
DB 122 PGSSPSVQCRSPRGKNIQGGKTLVSQLELDSDGTMCTVLOKQKVEFKIDIVLAFOK 181
QY 207 ASSIVYKKEGEQVEFSPFLAFTVEKLTGSGELMWQERASSSKSMITFDLKNKEVSRYR 266
DB 182 ASSIVYKKEGEQVEFSPFLAFTVEKLTGSGELMWQERASSSKSMITFDLKNKEVSRYR 241
QY 267 TQDPKLOMGKKLPLHLTLPALPOYAGSGNLTALAEKTKGLHQBENVLVYMRATQLOK 326
DB 242 TQDPKLOMGKKLPLHLTLPALPOYAGSGNLTALAEKTKGLHQBENVLVYMRATQLOK 301
QY 327 TCEVWGPTSPKMLSLKLENKAVSKREKPVWVLANPEAGMOCCLSDSGOVLLESNIK 386
DB 302 TCEVWGPTSPKMLSLKLENKAVSKREKPVWVLANPEAGMOCCLSDSGOVLLESNIK 361
QY 387 LPTWSTPV 394
DB 362 LPTWSTPV 369

RESULT 3
RMMQ74
T-cell surface glycoprotein CD4 - rhesus macaque
N:Alternate names: T-cell surface antigen T4/Leu 3
C:Species: Macaca mulatta (rhesus macaque)

C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Jul-1999
R:Accession: C32722
R:Camertini, D.; Seed, B.
Cell 60, 747-754, 1990
A:Title: A CD4 domain important for HIV-mediated syncytium formation lies outside the v
A:Reference number: A32722; MUID:90182664; PMID:2107024
A:Accession: C32722
A:Molecule type: mRNA
A:Residues: 1-432 <CM>
A:Cross-references: GB:M31134
C:Comment: This protein is expressed on most thymocytes, on a subset of mature T-cells
C:Superfamily: T-cell surface glycoprotein CD4; immunoglobulin homology
F:1-432/Product: T-cell surface glycoprotein CD4 #status predicted <MAT>
F:1-371/Domains: extracellular #status predicted <EXT>
F:9-86/Domains: immunoglobulin homology <IM1>
F:111-161/Domains: immunoglobulin homology #status atypical <IM2>
F:180-293/Domains: immunoglobulin homology <IM3>
F:296-347/Domains: immunoglobulin homology <IM4>
F:372-395/Domains: transmembrane #status predicted <TM>
F:396-432/Domains: intracellular #status predicted <INT>
F:16-84,130-159,303-345/Disulfide bonds: #status predicted
F:271,300/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 50.6%; Score 1729; DB 1; Length 432;
Best Local Similarity 90.8%; Pred. No. 1.9e-94;
Matches 334; Conservative 16; Mismatches 18; Indels 0; Gaps 0;

QY 27 KVLGKKGDVLTCTASQKKSIOFHWKNSNQIKILGNQGSFLTKGSKLNDRAISRSL 86
DB 2 KVLGKKGDVLTCTASQKKSIOFHWKNSNQIKILGNQGSFLTKGSKLNDRAISRSL 61
QY 87 WDQGNFPLITIKNLKIDSDTYICEVEDQKEVOLLVGLTANSDTHLLOQSILTLTLESP 146
DB 62 WDQGNFPLITIKNLKIDSDTYICEVEDQKEVOLLVGLTANSDTHLLOQSILTLTLESP 121
QY 147 PGSSPSVQCRSPRGKNIQGGKTLVSQLELDSDGTMCTVLOKQKVEFKIDIVLAFOK 206
DB 122 PGSSPSVQCRSPRGKNIQGGKTLVSQLELDSDGTMCTVLOKQKVEFKIDIVLAFOK 181
QY 207 ASSIVYKKEGEQVEFSPFLAFTVEKLTGSGELMWQERASSSKSMITFDLKNKEVSRYR 266
DB 182 ASSIVYKKEGEQVEFSPFLAFTVEKLTGSGELMWQERASSSKSMITFDLKNKEVSRYR 241
QY 267 TQDPKLOMGKKLPLHLTLPALPOYAGSGNLTALAEKTKGLHQBENVLVYMRATQLOK 326
DB 242 TQDPKLOMGKKLPLHLTLPALPOYAGSGNLTALAEKTKGLHQBENVLVYMRATQLOK 301
QY 327 TCEVWGPTSPKMLSLKLENKAVSKREKPVWVLANPEAGMOCCLSDSGOVLLESNIK 386
DB 302 TCEVWGPTSPKMLSLKLENKAVSKREKPVWVLANPEAGMOCCLSDSGOVLLESNIK 361
QY 387 LPTWSTPV 394
DB 362 VPTWPTPV 369

RESULT 4
GHU
Ig gamma-1 chain C region - human
C:Species: Homo sapiens (man)
C:Date: 31-Jan-1981 #sequence_revision 18-Aug-1982 #text_change 16-Jul-1999
R:Accession: A93433; S36867; S3887; B90563; A90564; B91688; A91723; A02146
R:Ellison, J.W.; Berson, B.J.; Hood, L.E.
Nucleic Acids Res. 10, 4071-4079, 1982
A:Title: The nucleotide sequence of a human immunoglobulin C-gamma1 gene.
A:Reference number: A93433; MUID:82274238; PMID:6287432
A:Accession: A93433
A:Molecule type: DNA
A:Residues: 1-330 <EUL>
A:Cross-references: EMBL:217370
A>Note: this sequence has the Gln(17) allelic marker, 97-Lys, and the Gln(1) markers,
A:Note: Lys-330 is removed after translation

R,Harris, L.J.
 submitted to the EMBL Data Library, October 1992
 A/Reference number: S33904
 A/Accession: S36861
 A/Molecule type: DNA
 A/Residues: 2-330 <HAR>
 A/Cross-references: EMBL:Z17370
 R/Takahashi, N.; Ueda, S.; Obata, M.; Nikaido, T.; Nakai, S.; Honjo, T.
 Cell 29, 671-679, 1982
 A/Title: Structure of human immunoglobulin gamma genes: implications for evolution of a
 A/Reference number: S33887; MUID:83001943; PMID:6811139
 A/Accession: S33887
 A/Molecule type: DNA
 A/Residues: 88-113;235-330 <TAK>
 A/Cross-references: EMBL:Z17370
 R/Cunningham, B.A.; Rutishauser, U.; Gall, W.E.; Gottlieb, P.D.; Maxdai, M.J.; Edelman,
 Biochemistry 9, 3161-3170, 1970
 A/Title: The covalent structure of a human gammaG-immunoglobulin. VII. Amino acid sequen
 A/Reference number: A90563; MUID:71064024; PMID:5489771
 A/Accession: myeloma protein Eu
 A/Contents: B90563
 A/Molecule type: protein
 A/Residues: 1-96, 'R', 98-135 <CUN>
 A/Note: this sequence has the G1m(3) marker, 97-Arg
 R/Rutishauser, U.; Cunningham, B.A.; Bennett, C.; Konigsberg, W.H.; Edelman, G.M.
 Biochemistry 9, 3171-3181, 1970
 A/Title: The covalent structure of a human gammaG-immunoglobulin. VIII. Amino acid sequen
 A/Reference number: A90564; MUID:71064025; PMID:5530842
 A/Accession: Eu
 A/Contents: A90564
 A/Molecule type: protein
 A/Residues: 136-154, 'Q', 156-165, 'Q', 167-176, 'Q', 178-194, 'N', 196-197, 'D', 199-238, 'E', 240,
 A/Note: this sequence has the G1m(non-1) markers, 239-Glu and 241-Met
 R/Ponstingl, H.; Hilschmann, N.
 Hoppe-Seyler's Z. Physiol. Chem. 357, 1571-1604, 1976
 A/Title: Die Primarstruktur eines monoklonalen IgG1-Immunglobulins (Myelomprotein Nie),
 Igen Primarstruktur.
 A/Reference number: A91668; MUID:77070269; PMID:826475
 A/Accession: myeloma protein Nie
 A/Contents: B91668
 A/Molecule type: protein
 A/Residues: 1-34, 'Q', 36-96, 'K', 98-115, 'Q', 117-197, 'D', 199-238, 'D', 240, 'L', 242-268, 'E', 27
 A/Note: this sequence has the G1m(17) and G1m(1) markers
 R/Schmidt, W.E.; Jung, H.D.; Palm, W.; Hilschmann, N.
 Hoppe-Seyler's Z. Physiol. Chem. 364, 713-747, 1983
 A/Title: Die Primarstruktur des Kristallisierten monoklonalen Immunglobulins IgG1 KOI
 A/Reference number: A91723; MUID:83289131; PMID:6884994
 A/Contents: myeloma protein KOI; disulfide bonds
 A/Accession: A91723
 A/Molecule type: protein
 A/Residues: 1-96, 'R', 98-197, 'D', 199-238, 'E', 240, 'W', 242-266, 'D', 268-271, 'D', 273-330 <SCH
 A/Note: this sequence has the G1m(3) and G1m(non-1) markers
 R/Gall, W.E.; Edelman, G.M.
 Biochemistry 9, 3188-3196, 1970
 A/Title: The covalent structure of a human gammaG-immunoglobulin. X. Intrachain disulfid
 A/Reference number: A90565; MUID:71064027; PMID:4923144
 A/Contents: annotation; disulfide bonds
 R/Drexler, L.; Schwarz, J.; Reichel, W.; Hilschmann, N.
 Hoppe-Seyler's Z. Physiol. Chem. 357, 1515-1540, 1976
 A/Title: Rule of antibody structure. The primary structure of monoclonal IgG1 immunoglob
 endomide cleavage products, and the disulfide bridges.
 A/Reference number: A91667; MUID:77070267; PMID:1002129
 A/Contents: annotation; disulfide bonds
 C/Genetics:
 A/Gene: GDB:IGHG1
 A/Cross-references: GDB:120085; OMIM:147100
 A/Map position: 14q32.33-14q32.33
 A/Intons: 99/1, 114/1, 224/1
 C/Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kap
 hain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into la
 C/Superfamily: immunoglobulin C region; immunoglobulin homology
 C/Keywords: duplication; glycoprotein; heterotetramer; immunoglobulin
 F/20-85/Domain: immunoglobulin homology <IM1>

F/137-206/Domain: immunoglobulin homology <IM2>
 F/243-310/Domain: immunoglobulin homology <IM3>
 F/27-83,144-204,250-308/Disulfide bonds: #status experimental
 F/103/Disulfide bonds: interchain (to light chain) #status experimental
 F/109,112/Disulfide bonds: interchain (to heavy chain) #status experimental
 F/180/Binding site: carbohydrate (Aen) (covalent) #status experimental
 Query Match 35.5%; Score 121.5; DB 1; Length 330;
 Best Local Similarity 97.8%; Pred. No. 3.6e-64;
 Matches 224; Conservative 0; Mismatches 0; Indels 5; Gaps 1;
 400 EPKSCDKHTTC-----PELLGSPVFLFPKPKDTLMISRTPEVTCVVVDVSHEDPEVKF 454
 99 EPKSCDKHTTCPPCPAPPELLGSPVFLFPKPKDTLMISRTPEVTCVVVDVSHEDPEVKF 158
 455 NMVYDGEVHNKTKPREQVNSTYRVSVLTVLHODWLNKGYKCKVSNKALPAPIEKT 514
 159 NMVYDGEVHNKTKPREQVNSTYRVSVLTVLHODWLNKGYKCKVSNKALPAPIEKT 218
 515 ISKAGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEESGQPENNYKTP 574
 219 ISKAGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEESGQPENNYKTP 278
 575 PVLDSGSEFFLYSKLTVDSRWQGNVFSQVMEALHNHYTQKSLSLSPG 625
 279 PVLDSGSEFFLYSKLTVDSRWQGNVFSQVMEALHNHYTQKSLSLSPG 329
 RESULT 5
 569339
 Ig heavy chain V region precursor - human
 C/Species: Homo sapiens (man)
 C/Date: 19-Mar-1997 #sequence revision 19-Mar-1997 #ext_change 01-Dec-2000
 C/Accession: S69339; S72664
 R/Khamlichi, A.A.; Aucouturier, P.; Preud'homme, J.L.; Cogne, M.
 Eur. J. Biochem. 229, 54-60, 1995
 A/Title: Structure of abnormal heavy chains in human heavy-chain-deposition disease.
 A/Reference number: S69339; MUID:95262687; PMID:7744049
 A/Accession: S69339
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-374 <KHA>
 A/Cross-references: EMBL:X81695
 R/Khamlichi, A.A.
 submitted to the EMBL Data Library, September 1994
 A/Reference number: S72664
 A/Accession: S72664
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-140, 'C', 142-374 <KH2>
 A/Cross-references: EMBL:X81695
 C/Superfamily: immunoglobulin C region; immunoglobulin homology
 Query Match 35.3%; Score 1205.5; DB 2; Length 374;
 Best Local Similarity 97.0%; Pred. No. 9.6e-64;
 Matches 224; Conservative 2; Mismatches 0; Indels 5; Gaps 1;
 400 EPKSCDKHTTC-----PELLGSPVFLFPKPKDTLMISRTPEVTCVVVDVSHEDPEVKF 454
 143 EPKSCDKHTTCPPCPAPPELLGSPVFLFPKPKDTLMISRTPEVTCVVVDVSHEDPEVKF 202
 455 NMVYDGEVHNKTKPREQVNSTYRVSVLTVLHODWLNKGYKCKVSNKALPAPIEKT 514
 203 NMVYDGEVHNKTKPREQVNSTYRVSVLTVLHODWLNKGYKCKVSNKALPAPIEKT 262
 515 ISKAGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEESGQPENNYKTP 574
 263 ISKAGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEESGQPENNYKTP 322
 575 PVLDSGSEFFLYSKLTVDSRWQGNVFSQVMEALHNHYTQKSLSLSPG 625
 323 PVLDSGSEFFLYSKLTVDSRWQGNVFSQVMEALHNHYTQKSLSLSPG 373

RESULT 6

S31866
Ig gamma-1 chain C region - synthetic

C/Species: synthetic
A/Note: Homo sapiens (man) gene engineered and expressed in Escherichia coli
C/Date: 06-Jan-1995 #sequence_revision 17-Mar-1997 #text_change 19-May-2000
C/Accession: S31866

R/Flpula, D.

submitted to the EMBL Data Library, February 1993

A/Description: Screening method for protein-protein interactions of cloned gene products.

A/Reference number: S31866

A/Accession: S31866

A/Molecule type: mRNA

A/Residues: 1-255 <Flp>

A/Cross-references: EMBL:X70421; NID:g33068; PIDN:CAA4866.1; PID:g33069

C/Keywords: immunoglobulin

F/1-22/Region: Escherichia coli outer membrane protein A precursor

F/23-255/Region: human Ig gamma-1 chain C region

Query Match 35.3%; Score 1203.5; DB 4; Length 255;
Best Local Similarity 97.4%; Pred. No. 7.7e-64;

Matches 225; Conservative 0; Mismatches 1; Indels 5; Gaps 1;

Qy 400 EPKSCDKHTHC-----PELLGSPSVFLPPPKKDTLMISRTPEVTCVVVDVSHEDPEVKF 454
Dy 24 ESKSCDKHTHCPCAPPELLGSPSVFLPPPKKDTLMISRTPEVTCVVVDVSHEDPEVKF 83

Qy 455 NMVYDGEVHNAAKTPREEOYNSTYRVSVLTVLHQMNLNKEVKCKVSNKALPAPIEKT 514
Dy 84 NMVYDGEVHNAAKTPREEOYNSTYRVSVLTVLHQMNLNKEVKCKVSNKALPAPIEKT 143

Qy 515 ISKAGGOREPOVYTLPPSRDELTKNOVSLCLVKGFPSPDIAVWESNGOPENNYKTP 574
Dy 144 ISKAGGOREPOVYTLPPSRDELTKNOVSLCLVKGFPSPDIAVWESNGOPENNYKTP 203

Qy 575 PVLDSGDFLYSKLTVDKSRMOQGNVSCVMHEALHNHTOKSLSPG 625
Dy 204 PVLDSGDFLYSKLTVDKSRMOQGNVSCVMHEALHNHTOKSLSPG 254

RESULT 7

PT0207

Ig gamma chain C region - chimpanzee

C/Species: Pan troglodytes (chimpanzee)

C/Date: 23-Nov-1991 #sequence_revision 23-Nov-1991 #text_change 16-Jul-1999

C/Accession: PT0207

R/Enrlich, P.H.; Moustafa, Z.A.; Oestberg, L.

Mol. Immunol. 28, 319-322, 1991

A/Title: Nucleotide sequence of chimpanzee Fc and hinge regions.

A/Reference number: PT0207; MUID:91287716; PMID:2062315

A/Accession: PT0207

A/Molecule type: mRNA

A/Residues: 1-234 <EHR>

C/Suprafamily: immunoglobulin C region; immunoglobulin homology

C/Keywords: immunoglobulin

F/48-117/Domain: immunoglobulin homology <IMM>

Query Match 34.3%; Score 1171.5; DB 2; Length 234;
Best Local Similarity 96.9%; Pred. No. 5.3e-62;

Matches 218; Conservative 1; Mismatches 1; Indels 5; Gaps 1;

Qy 400 EPKSCDKHTHC-----PELLGSPSVFLPPPKKDTLMISRTPEVTCVVVDVSHEDPEVKF 454
Dy 10 EPKSCDTHTCPCCAPPELLGSPSVFLPPPKKDTLMISRTPEVTCVVVDVSHEDPEVKF 69

Qy 455 NMVYDGEVHNAAKTPREEOYNSTYRVSVLTVLHQMNLNKEVKCKVSNKALPAPIEKT 514
Dy 70 NMVYDGEVHNAAKTPREEOYNSTYRVSVLTVLHQMNLNKEVKCKVSNKALPAPIEKT 129

Qy 515 ISKAGGOREPOVYTLPPSRDELTKNOVSLCLVKGFPSPDIAVWESNGOPENNYKTP 574
Dy 130 ISKAGGOREPOVYTLPPSRDELTKNOVSLCLVKGFPSPDIAVWESNGOPENNYKTP 189

Qy 575 PVLDSGDFLYSKLTVDKSRMOQGNVSCVMHEALHNHTOKS 619
Dy 190 PVLDSGDFLYSKLTVDKSRMOQGNVSCVMHEALHNHTOKS 234

RESULT 8

A23511

Ig gamma-3 chain C region (allotype G3m(b)) - human

C/Species: Homo sapiens (man)

C/Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change 23-Jul-1999

C/Accession: A23511

R/Huck, S.; Fort, P.; Crawford, D.H.; Lefranc, M.P.; Lefranc, G.

Nucleic Acids Res. 14, 1779-1789, 1986

A/Title: Sequence of a human immunoglobulin gamma 3 heavy chain constant region gene: c

A/Reference number: A23511; MUID:86148507; PMID:3081877

A/Accession: A23511

A/Molecule type: DNA

A/Residues: 1-377 <HUC>

A/Cross-references: GB:X03604; GB:M2958; NID:g33070; PIDN:CAA27268.1; PID:g577056

C/Genetics:

A/Gene: IGHG3

A/Cross-references: GDB:119339; OMIM:147120

A/Map position: 14q32.33-14q32.33

A/Intons: 98/3; 115/3; 130/3; 145/3; 160/3; 270/3

C/Suprafamily: immunoglobulin C region; immunoglobulin homology

C/Keywords: immunoglobulin

F/20-85/Domain: immunoglobulin homology <IMM>

Query Match 34.1%; Score 1163; DB 2; Length 377;
Best Local Similarity 90.0%; Pred. No. 3.1e-61;

Matches 216; Conservative 8; Mismatches 10; Indels 6; Gaps 2;

Qy 392 TPVPCP-APBPKSCDKHTHC-----PELLGSPSVFLPPPKKDTLMISRTPEVTCVVVDV 445
Dy 137 TPVPCPCPCPKSCDTPPCPCAPPELLGSPSVFLPPPKKDTLMISRTPEVTCVVVDV 196

Qy 446 SHEDPEVKFMVYDGEVHNAAKTPREEOYNSTYRVSVLTVLHQMNLNKEVKCKVSNK 505
Dy 197 SHEDPEVKFMVYDGEVHNAAKTPREEOYNSTYRVSVLTVLHQMNLNKEVKCKVSNK 256

Qy 506 ALPAPIEKTISKAGGOREPOVYTLPPSRDELTKNOVSLCLVKGFPSPDIAVWESNGO 565
Dy 257 ALPAPIEKTISKAGGOREPOVYTLPPSRDELTKNOVSLCLVKGFPSPDIAVWESNGO 316

Qy 566 PENNYKTPPVLDSGDFLYSKLTVDKSRMOQGNVSCVMHEALHNHTOKSLSPG 625
Dy 317 PENNYKTPPVLDSGDFLYSKLTVDKSRMOQGNVSCVMHEALHNHTOKSLSPG 376

RESULT 9

A60764

Ig gamma-3 chain C region, form LAT - human

C/Species: Homo sapiens (man)

C/Date: 14-May-1993 #sequence_revision 14-May-1993 #text_change 16-Jul-1999

C/Accession: A60764

R/Huck, S.; Lefranc, G.; Lefranc, M.P.

Immunogenetics 30, 250-257, 1989

A/Title: A human immunoglobulin IGHG3 allele (Gmb0, b1, c3, c5, u) with an IGHG4 conver

A/Reference number: A60764; MUID:90007613; PMID:2571587

A/Accession: A60764

A/Status: preliminary

A/Molecule type: DNA

C/Suprafamily: immunoglobulin C region; immunoglobulin homology

C/Keywords: immunoglobulin

F/20-85/Domain: immunoglobulin homology <IMM>

Query Match 34.0%; Score 1161; DB 2; Length 377;
Best Local Similarity 90.0%; Pred. No. 4e-61;

Matches 216; Conservative 8; Mismatches 10; Indels 6; Gaps 2;

Qy 392 TPVPCP-APBPKSCDKHTHC-----PELLGSPSVFLPPPKKDTLMISRTPEVTCVVVDV 445

```

Db      137  TEPPCRCRCEBPSCDTPPCRCRCPABELLGGEVFLFPKPKDTLMISTPVTCTVAVV 196
      446  SHEDPEVKFNWYDGVGVHNAKTKPREEQYNSTRVSVLTVLHODMLNGEKYCKVSNK 505
      197  SHEDPEVQFKWYDGVGVHNAKTKPREEQYNSTRVSVLTVLHODMLNGEKYCKVSNK 256
      506  ALPAPLEKITSKAKGQPREQVYTLTPSRDELTKNOVSLTCLVKGYFSPSDIAVEMESNGQ 565
      257  ALPAPLEKITSKAKGQPREQVYTLTPSRDELTKNOVSLTCLVKGYFSPSDIAVEMESNGQ 316
      566  PENNYKTPPVLDSDFLYSKLTVDKSRMOQGNFSCSVMEALAHNYTQKSLSLSPG 625
      317  PENNYKTPPVLDSDFLYSKLTVDKSRMOQGNFSCSVMEALAHNYTQKSLSLSPG 376

```

RESULT 10

```

A:Accession: A46254
CD4 precursor - rabbit
C/Spectrum: Oryctolagus cuniculus (domestic rabbit)
C/Date: 21-Sep-1993 #sequence, revision 18-Nov-1994 #text, change 21-Jan-2000
C/Accession: A46254
R/Hague, B.F.; Sawaadikosol, S.; Brown, T.J.; Lee, K.; Becker, D.P.; Kindt, T.J.
Proc. Natl. Acad. Sci. U.S.A. 89, 7963-7967, 1992
A>Title: CD4 and its role in infection of rabbit cell lines by human immunodeficiency vi
A/Reference number: A46254; PMID:92390370; PMID:1518821
A/Accession: A46254
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-459 <HAG>
A/Note: reference: GB:W92840; NID:9164871; PIN:AAA31198.1; PID:9164872
C/Superfamily: T-cell surface glycoprotein CD4; immunoglobulin homology
F/322-372/Domain: immunoglobulin homology <ITM>

```

```

Query Match      33.7%; Score 1150.5; DB 2; Length 459;
Best Local Similarity 57.2%; Pred. No. 2, 1e-60;
Matches 241; Conservative 65; Mismatches 90; Indels 25; Gaps 6;

```

```

      1  MNRGVFRLHLVLAALPAATQGNKVLGKKGDTVELTCTASQKKSIOFHKNKNQIK 60
      1  MNRVIFQCLLVLPALPLPAATMGKTVVKGAGATVELPCQSSQKNSVFNKHNQVYK 60
      61  ILGNQSSSSSSFWLKGNSPLSNVESKKNWMDGSPFLVTKDLMDSGTYICEVGDCKV 120
      61  ILGNQSSSSSSFWLKGNSPLSNVESKKNWMDGSPFLVTKDLMDSGTYICEVGDCKV 120
      117  EVOLLVFGLTANSDFHLLQGSITLTLSPSPSSPVQCSPRGKNIQGGKTLVSQLEL 176
      121  EVLLVFLRLTANTRNLTHQOSITLTLSEBSVGSPSVQWSPENKIETGPTCSMKRL 180
      177  QDSGTWTCV-LQNKQKVEFKIDIVLAFQKASSIYKKEGEVSPFLAFTVEKLTGS 235
      181  QDSGTWTCV-LQNKQKVEFKIDIVLAFQKASSIYKKEGEVSPFLAFTVEKLTGS 235
      236  GELMWAQERASSSSKSWTTPPLKNKEVSVKVTODPKQMKKPLHLTLPLQALPQVAGS 255
      239  GELMWAQERASSSSKSWTTPPLKNKEVSVKVTODPKQMKKPLHLTLPLQALPQVAGS 255
      239  GELMWAQERASSSSKSWTTPPLKNKEVSVKVTODPKQMKKPLHLTLPLQALPQVAGS 255
      239  GELMWAQERASSSSKSWTTPPLKNKEVSVKVTODPKQMKKPLHLTLPLQALPQVAGS 255
      296  NLTLALAKTKGLHGVNVLVVRATOLQKVLTCVEMGPTSPKMLSLKLENKEAKYSKRE 355
      299  NLTLALAKTKGLHGVNVLVVRATOLQKVLTCVEMGPTSPKMLSLKLENKEAKYSKRE 355
      356  KPVAVINPEAGMMQCLLSDSGVLLSESNIVLPTWSTPVPAPPEPKSCDKTHTCEPLL 415
      356  KPVAVINPEAGMMQCLLSDSGVLLSESNIVLPTWSTPVPAPPEPKSCDKTHTCEPLL 415
      416  G 416
      401  G 401

```

RESULT 11

```

G3HUM1
Ig gamma-3 heavy chain disease proteins - human
C/Spectrum: Homo sapiens (man)
C/Date: 31-Dec-1979 #sequence, revision 23-Oct-1991 #text, change 16-Jul-1999
C/Accession: A90442; A92219; A90198; A93915; A02149
R/Frangione, B.; Rosenwasser, E.; Prelli, F.; Franklin, E.C.
Biochemistry 19, 4304-4308, 1980
A>Title: Primary structure of human gamma3 immunoglobulin deletion mutant: gamma3 heavy-
A/Reference number: A90442; PMID:81021548; PMID:6774747
A/Accession: A90442
A/Molecule type: protein
A/Residues: 1-289 <FRA>
A/Note: the molecule is a dimer linked by 12 disulfide bonds; it has an extra interchain
A/Note: this protein lacks most of the V region and all of the CH1 region. Residue 12 c
A/Note: the sequence of residues 42-76 was taken from the reference that follows
J. Michaelson, T.E.; Frangione, B.; Franklin, E.C.
J. Biol. Chem. 252, 883-889, 1977
A>Title: Primary structure of the 'hinge' region of human IgG3. Probable quadruplication
A/Reference number: A92219; PMID:77118561; PMID:402363
A/Accession: A92219
A/Status: normal gamma-3 chains, sequence corresponding to residues 12-97 of protein 1
A/Molecule type: protein
A/Residues: 12-97 <MIC>
A/Note: the hinge region in gamma-3 chains is about four times as long as in other gamma-
A/Note: cysteines at positions 24, 27, 33, 39, 42, 48, 54, 57, 63, 69, and 72 form inter
R/Mollenhede, Todel, C.; Frangione, B.; Prelli, F.; Franklin, E.C.
Biochem. Biophys. Res. Commun. 71, 907-914, 1976
A>Title: The amino acid sequence of "heavy chain disease" protein ZUC. Structure of the
A/Reference number: A90198; PMID:77021516; PMID:923945
A/Accession: A90198
A/Status: heavy chain disease protein ZUC, partial sequence corresponding to residues
A/Molecule type: protein
A/Residues: 59-125, 'EB', 128-226, 228-289 <MOL>
A/Note: this protein lacks most of the V region, all of the CH1 region, and part of the
R/Alexander, A.; Steinmetz, M.; Barilic, D.; Frangione, B.; Franklin, E.C.; Hood, L.
Proc. Natl. Acad. Sci. U.S.A. 79, 3260-3264, 1982
A>Title: gamma heavy chain disease in man: cDNA sequence supports partial gene deletion
A/Reference number: A93915; PMID:82247835; PMID:6808505
A/Accession: A93915
A/Status: heavy chain disease protein Omni
A/Molecule type: mRNA
A/Residues: 12-70, 72-114, 116-125, 'E', 127-133, 'L', 135-136, 'E', 138, 'Y', 140-154, 'D', 156-15
A/Note: a carboxyl-terminal lys is removed posttranslationally
A/Note: this sequence may represent an allelic form or another gamma chain subclass
C/Comment: The heavy chain disease protein wis is shown.
C/Accession: A93915
A/Status: heavy chain disease protein Omni
A/Map position: 14q32.33-14q32.33
C/Superfamily: immunoglobulin C region; immunoglobulin homology
F/203-270/Domain: immunoglobulin homology <ITM>
F/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental
F/6,140/Binding site: carbohydrate (Asn) (covalent) #status experimental

```

```

Query Match      33.5%; Score 1143; DB 1; Length 289;
Best Local Similarity 87.9%; Pred. No. 3, 3e-60;
Matches 211; Conservative 13; Mismatches 10; Indels 6; Gaps 2;

```

```

      392  TTPPCRCRCEBPSCDTPPCRCRCPABELLGGEVFLFPKPKDTLMISTPVTCTVAVV 445
      50  TTPPCRCRCEBPSCDTPPCRCRCPABELLGGEVFLFPKPKDTLMISTPVTCTVAVV 109
      446  SHEDPEVKFNWYDGVGVHNAKTKPREEQYNSTRVSVLTVLHODMLNGEKYCKVSNK 505
      110  SHEDPEVQFKWYDGVGVHNAKTKPREEQYNSTRVSVLTVLHODMLNGEKYCKVSNK 169
      506  ALPAPLEKITSKAKGQPREQVYTLTPSRDELTKNOVSLTCLVKGYFSPSDIAVEMESNGQ 565
      170  ALPAPLEKITSKAKGQPREQVYTLTPSRDELTKNOVSLTCLVKGYFSPSDIAVEMESNGQ 229

```

Oy 566 PENNYKTTTPPLDSDGSFFLYSKLTJTDKSRWQGNVSCSVHHEALHHHTQSLSLSPG 623
 ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 230 PENNTTTPMLDSDGSFFLYSKLTJTDKSRWQGNIFCSVHHEALNHRFTQSLSLSPG 289

RESULT 12

IG gamma-2 chain C region - human
C:Species: Homo sapiens (man)
C:Date: 30-Apr-1991 #sequence_revision 13-Jun-1993 #text_change 21-Jul-2000
C:Accession: A93906; A92809; A90752; A93132; A02148
R:Ellison, J.; Hood, L.
Proc. Natl. Acad. Sci. U.S.A. 79, 1984-1988, 1982
A:Title: Linkage and sequence homology of two human immunoglobulin gamma heavy chain c
A:Reference number: A93906; MUID:82197621; PMID:6804948
A:Accession: A93906
A:Molecule type: DNA
A:Residues: 1-326 <ELL>
A:Cross-references: GB:V00554; GB:J00230; NID:932759; PIDN:CA858438.1; FID:g6066056
A:Note: Lys-326 is probably removed posttranslationally
R:Wang, A.C.; Tung, E.; Fudenberg, H.H.
J. Immunol. 125, 1048-1054, 1980
A:Title: The primary structure of a human IgG2 heavy chain: genetic, evolutionary, and f
A:Reference number: A92809; MUID:81007873; PMID:6774012
A:Accession: A92809
A:Molecule type: protein
A:Residues: 1-19, 'Q', 21-57, 'Z', 59, 'A', 61-193, 'D', 195-325 <MAN>
A:Note: Trp-156 is at or near the complement-binding site
R:Connell, G.E.; Parr, D.M.; Hofmann, T.
Can. J. Biochem. 57, 758-767, 1979
A:Title: The amino acid sequences of the three heavy chain constant region domains of a
A:Reference number: A90752; MUID:8000357; PMID:113060
A:Accession: A90752
A:Molecule type: protein
A:Residues: 1-24, 'E', 26-57, 'EV', 60-85, 132-171, 'ZZZ', 175, 'B', 177-193, 'D', 195-196, 'Q', 198
A:Note: this sequence has since been revised
R:Hofmann, T.; Parr, D.M.
Mol. Immunol. 16, 923-925, 1979
A:Title: A note on the amino acid sequence of residues 381-391 of human immunoglobulin g
A:Reference number: A93132; MUID:80114419; PMID:118920
A:Accession: A93132
A:Molecule type: protein
A:Residues: 238-275 <HOF>
R:Hofmann, T.; Parr, D.M.
submitted to the Atlas, March 1980
A:Reference number: A94591
A:Contents: annotation; Zie, revisions to residues 25, 59, 60, and 264-268
A:Note: the revised sequence differs from that shown in having 60-Ala and in the amidat
ned
R:Malstein, C.; Frangione, B.
Biochem. J. 121, 217-225, 1971
A:Title: Disulfide bridges of the heavy chain of human immunoglobulin G2.
A:Reference number: A90253; MUID:72033500; PMID:4640472
A:Contents: annotation; myeloma protein 8s, disulfide bonds
R:Frangione, B.; Malstein, C.; Plink, J.R.L.
Nature 221, 145-148, 1969
A:Title: Structural studies of immunoglobulin G.
A:Reference number: A93157; MUID:69064124; PMID:5782707
A:Contents: annotation; Sa, disulfide bonds
C:Genetics:
A:Gene: GDB:IGHG2
A:Cross-references: GDB:119338; OMIM:147110
A:Map position: 14q32.33-14q32.33
C:Complex: An immunoglobulin heterotetramer subunit consists of two identical light (K)
chain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into la
C:Superfamily: immunoglobulin C region; immunoglobulin homology
C:Keywords: duplication; glycoprotein; heterotetramer; immunoglobulin
E:20-85/Domain: immunoglobulin homology <IM1>
F:133-202/Domain: immunoglobulin homology <IM2>
F:239-306/Domain: immunoglobulin homology <IM3>

F.14/Disulfide bonds: interchain (to light chain) #status experimental
F.27-83,140-200,246-304/Disulfide bonds: #status experimental
F.102,103,106,109-150/Disulfide bonds: interchain (to heavy chain) #status experimental
F.176/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match	32.7%;	Score 1115.5;	DB 1;	Length 326;
Best Local Similarity	89.2%;	Pred. No. 1.6e-58;		
Matches 206;	Conservative 8;	Mismatches 4;	Indels 13;	Gaps 1.

```

QY 395 PCAPAEKSCDKHTHCPELGLGSPVSLFPPKPKOTLMSLRPELVGVAVDVSHDEPEKFF 45
Db 108 PCAPRP-----VAGPSVFLFPPKPKOTLMSLRPELVGVAVDVSHDEPEVQF 15
QY 455 NMVYDGEVYHNAKTKREBEQYNSTYRVSVLTVLHQDLNGKEYKCKVSNKALPAPIEKT 51
Db 155 NMVYDGEVYHNAKTKREBEQYNSTPRVSVLTVLVHQDLNGKEYKCKVSNKGLPAPIEKT 21
QY 515 ISRAKGGPRPRQYVTLPRSRDELTKQVSLTGLVYGFPPSDIAYEMESNGCPENNNYKTP 57
Db 215 ISKTKGGPRPRQYVTLPRSRREETKQVSLTGLVYGFPPSDIAYEMESNGCPENNNYKTP 27
QY 575 PVLDDSDSPFLYISKLTIVDKSRRMOQGVFSCSWMEALNNHTQKSLSLSPG 625
Db 275 PVLDDSDSPFLYISKLTIVDKSRRMOQGVFSCSWMEALNNHTQKSLSLSPG 325

```

RESULT 13

```

Ig gamma-4 chain C region - human
C.Species: Homo sapiens (man)
C.Date: 02-Apr-1988 #sequence_revision 02-Apr-1992 #text_change 16-Jul-1999
C.Accession: A90933; A90249; A02150
R.Ellison, J.; Buxbaum, J.; Hood, L.
DNA 1, 11-18, 1981
A.Title: Nucleotide sequence of a human immunoglobulin C-gamma4 gene.
A.Reference number: A90933; MUID:83157104; PMID:6299662
A.Accession: A90933
A.Molecule type: DNA
A.Residues: 1-327 <ELL>
A.Note: The sequence was determined from the germ-line gene
R.Pink, J.R.L.; Buttery, S.H.; De Vries, G.M.; Mlstein, C.
Biochem. J. 117, 33-47, 1970
A.Title: Human immunoglobulin subclasses. Partial amino acid sequence of the constant
A.Reference number: A90249; MUID:70207560; PMID:4192699
A.Accession: A90249
A.Molecule type: Protein
A.Residues: 1-30; 81-326 <PIN>
C.Genetics:
A.Gene: GDB:IGHG4
A.Cross-references: GDB:119340; OMIM:147130
A.Map position: 14q32.33-14q32.33
A.Introns: 99/1; 111/1; 221/1
C.Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kappa)
chain disulfide bonds. In some cases, such as IGA and IGH, the subunits associate into 1
C.Superfamily: immunoglobulin C region; immunoglobulin homology
C.Keywords: duplication; glycoprotein; heterotetramer; immunoglobulin
F:20-85/Domain: immunoglobulin homology <IM1>
F:99-110/Region: hinge
F:134-203/Domain: immunoglobulin homology <IM2>
F:140-307/Domain: immunoglobulin homology <IM3>
F:14/Disulfide bonds: interchain (to light chain) #status experimental
F:27-83; 141-201; 247-305/Disulfide bonds: #status predicted
F:106; 109/Disulfide bonds: interchain (to heavy chain) #status experimental
F:117/Binding site: carboxylate (asn) (covalent) #status predicted

Query Match      32.4%   Score 1105.5; DB 1; Length 327;
      Beat Local Similarity 89.6%; Pred. No. 6.2e-58;
      Matches 207; Conservative 8; Mismatches 7; Indels 9; Gaps 2;

395 PCPAPRPSCCKTHCCPELLGGPSVFLPPPKYKDTLMSRPPEYVCVVVDVSHEDPEYKF 454
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      105 PCP-----SC-----PAPFLLGSPSVFLPPPKYKDTLMSRPPEYVCVVVDVSHEDPEYVF 155

```

Query Match	32.4%	Score 1105.5	DB 1	length 327
Best Local Similarity	89.6%	Pred. No. 6.2e-58		
Matches 207; Conservative	8;	Mismatches 7;	Indels 9;	Gaps 2

QY 395 PCPAPDPKSCDKTHICPELLGGPSVFLFPPPKDTLMISRPPEVTCVVDVSDHEDEPKVF 455
 105 PCP-----SC-----PAPEFLGGPSVFLFPPPKDTLMISRPPEVTCVVDVSDQEDPEPKVF 155
 Db

QY 455 NWVVDGEVHNATKPREOYNSTYRVSVLTVLHODMLNGKEYCKKCNKALPAPIEXT 514
 |||||
 DB 156 NWVVDGEVHNATKPREOYNSTYRVSVLTVLHODMLNGKEYCKKCNKALPAPIEXT 215
 |||||
 QY 515 ISKAKQPRPQYVTLPPSRDELTKNQVSLTCLVKGFPSDIAVEMESNQGPENNYKTP 574
 |||||
 DB 216 ISKAKQPRPQYVTLPPSRDELTKNQVSLTCLVKGFPSDIAVEMESNQGPENNYKTP 275
 |||||
 QY 575 PVLDSDGSFPLYSKLTVDKSRKNOQGVFSCSVHMEALHNYTOKSLSLSPG 625
 |||||
 DB 276 PVLDSDGSFPLYSKLTVDKSRKNOQGVFSCSVHMEALHNYTOKSLSLSPG 326
 |||||

RESULT 14

S30193
 T-cell surface glycoprotein CD4 - dog
 C/Species: Canis lupus familiaris (dog)
 C/Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 21-Jul-2000
 C/Accession: S30193
 R/Mide, K.F.; Conner, G.E.; Miner, D.H.; Alejandro, R.
 Biochim. Biophys. Acta 1172, 315-318, 1993
 A/Title: Primary structure of the canine CD4 antigen.
 A/Reference number: S30193; MID:93192324; PMID:7916632
 A/Accession: S30193
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-432 <MIL>
 A/Cross-references: EMBL:X68565; NID:9288652; PIDN:CA837664.1; PID:94467377
 C/Superfamily: T-cell surface glycoprotein CD4; immunoglobulin homology
 C/Keywords: glycoprotein
 F/202-311/Domain: immunoglobulin homology <IM>

Query Match 32.2%; Score 1099; DB 2; Length 432;
 Best Local Similarity 57.7%; Pred. No. 2, 1e-57;
 Matches 226; Conservative 62; Mismatches 86; Indels 18; Gaps 6;

QY 12 LVVLQALLPAATQGNKVVLGKGDVTELTCTASOKSIQPHMKNSQIKILNGQSFPLK 71
 |||||
 DB 1 LMLQVWLPVATPVREVLAKAGAVELPCQTSQKNIHNMKDSMWQTLNGQSFVWY 60
 |||||
 QY 72 GPKLNDRADSRSLMDQGNFPLIKNLKIEDSDTYICEVEDQKEEVQLVFGILTA---- 127
 |||||
 DB 61 GSSRLKRVESKKNLMQGSFPLVTKDLVADSGIYFCDF-DKQVEVLLVFLTKMDS 119
 |||||
 QY 128 ----NSDTHLLQGQSTLTLEBPSSPSVQCRSPKGNKIQGGKTLVSQLELDQSGTW 182
 |||||
 DB 120 GSSSGSSNIRLLQGQQLTLTLENPSGSSPSVQWKGPKNGKGGQNLSTSWPELDQGGTW 179
 |||||
 QY 183 TCTVLONQKVEFKIDIVLAFOKASSIVYKKEGQVFPSPPLAFVTEKLTGSGELMWQA 242
 |||||
 DB 180 TCTISQSQKTVERNINVLAFQVNSVTFYARBDQVDFSPFLSFEDENLV--GELRWQA 237
 |||||
 QY 243 ERASSSKSWITPDLKNEVSVKRVTDOPKLQMGKULPLHLTPQALPOVAGSNLTLE 302
 |||||
 DB 238 QGASSSLTWISFTLENNKLSMKEMHAPLKLQMKESLPRLRTLLQVLSRVAGSGLITLNL- 296
 |||||
 QY 303 AKTGKLGHOEVNLVVMRATQLOKULTCVWGPSTSPKMLSLKENKEAKYSKREKPVWVLN 362
 |||||
 DB 297 AK-GTLVQEVNLVVMRANSSQNNMLTCEVLTSPBELTSLNLKEQAAKYSKQKQLVWVVD 355
 |||||
 QY 363 PEAGMOCILSDSGOVLVLSNINIKVLPWSPNV 394
 |||||
 DB 356 PEAGTMOCLLSDKDKVLLASSLVN---SSPV 383
 |||||

RESULT 15

A27449
 T-cell surface glycoprotein CD4 precursor - rat
 N/Alternate names: W3/25 antigen
 C/Species: Rattus norvegicus (Norway rat)
 C/Date: 21-May-1988 #sequence_revision 21-May-1988 #text_change 21-Jan-2000
 C/Accession: A27449; A35433
 R/Clark, S.J.; Jefferies, W.A.; Barclay, A.N.; Gagnon, J.; Williams, A.F.

Proc. Natl. Acad. Sci. U.S.A. 84, 1649-1653, 1987
 A/Title: Peptide and nucleotide sequences of rat CD4 (W3/25) antigen: evidence for deriv
 A/Reference number: A27449; MID:87175535; PMID:3104900
 A/Accession: A27449
 A/Molecule type: mRNA
 A/Residues: 1-457 <CLA>
 A/Cross-references: GB:M15768; NID:9203387; PIDN:AAA40901.1; PID:9203388
 U. Davis, S.J.; Ward, H.A.; Pukavec, M.J.; Willie, A.C.; Williams, A.F.; Barclay, A.N.
 J. Biol. Chem. 265, 10410-10418, 1990
 A/Title: High level expression in Chinese hamster ovary cells of soluble forms of CD4 T
 A/Reference number: A35433; MID:90285164; PMID:2113054
 A/Contents: annotation
 C/Superfamily: T-cell surface glycoprotein CD4; immunoglobulin homology
 C/Keywords: glycoprotein; membrane protein; surface antigen
 F/219-300/Domain: immunoglobulin homology <IM>

Query Match 29.3%; Score 1000.5; DB 2; Length 457;
 Best Local Similarity 50.1%; Pred. No. 1, 4e-51;
 Matches 213; Conservative 64; Mismatches 127; Indels 21; Gaps 6;

QY 1 NNRGVPFRLL--LVVLQALLPAATQGNKVVLGKGDVTELTCTASOKSIQPHMKNSQ 58
 |||||
 DB 1 MCRGSPFRLLDLILLQSLKLVVTOGKTVLGEKGSALPCESTSRSSASAPWMSDQ 60
 |||||
 QY 59 IKILNGQSFPLTKGSPKLNDRADSRSLMDQGNFPLIKNLKIEDSDTYICEVEDQKEEV 118
 |||||
 DB 61 KTLIGYKKNLTKGSIELYSRPSRKNAMEGSPFLIKNLKIEDSDTYICEVEDQKEEV 120
 |||||
 QY 119 QLLVPELTANSPTLHLLQGSSTLTLES--PPGSSPSVQCRSPKGNKIQGGKTLVSQLELDQ 177
 |||||
 DB 121 ELWVFRVTNPGTRILLQGSSTLTLDNSPKVSDPTECHGKSNIVKDSKAFSTHSLRIQ 180
 |||||
 QY 178 DSGTWTCTYLQOKKVEFKIDIVLAFOKASSIVYKKEGQVFPSPPLAFVTEKLTGSGE 237
 |||||
 DB 181 DSGINWCTYTLQOKGSPFMKLSVLFQASTSTAYKSEBSAEFPSPMLGSESL--QGE 238
 |||||
 QY 238 LWMQERASSSKSWITPDLKNEVSVKRVTDOPKLQMGKULPLHLTPQALPOVAGSNLT 297
 |||||
 DB 239 LMKAKASQSSQWITPFLSKQKVSQKSTSNPKQLSTPLTLOIPQVSLQFAGSNLT 298
 |||||
 QY 298 TLAEAKTGKLGHOEVNLVVMRATQLOK--LTCEWGPSTSPKMLSLKENKEAKYSKREK 356
 |||||
 DB 299 TLTLTD--RGLVQEVNLVVMKVTQPONSNTLTCEWGPSTSPKMLSLKENKEARVGROR 356
 |||||
 QY 357 PPMVNLPEAGMOCILSDSGOVLVLSNINIKVLPWSPNVCPAPBEPSCDKTHCPDLLG 416
 |||||
 DB 357 VLQVAPPEAGVMOCLLSEGEVWMDSKIQVL-----SKGLNQTFLLAVLGS 403
 |||||
 QY 417 PSVFL 421
 |||||
 DB 404 AFSFL 408
 |||||

RESULT 16

RMMST4
 T-cell surface glycoprotein CD4 precursor - mouse
 N/Alternate names: T-cell differentiation antigen L3T4; T-cell surface antigen T4/Len 3
 C/Species: Mus musculus (house mouse)
 C/Date: 30-Jun-1987 #sequence_revision 30-Jun-1987 #text_change 16-Jul-1999
 C/Accession: A02110; A26038; A39893; A39953; I54564; I6018; A47642
 R/Pourviel, B.; Gorman, S.D.; Field, E.H.; Hunkapiller, T.; Parnes, J.R.
 Science 234, 610-614, 1986
 A/Title: Isolation and sequence of L3T4 complementary DNA clones: expression in T cells
 A/Reference number: A02110; MID:87018845; PMID:3094146
 A/Accession: A02110
 A/Molecule type: mRNA
 A/Residues: 1-457 <TOU>
 A/Cross-references: GB:M13816; NID:9192070; PIDN:AAA37267.1; PID:9309112
 R/Ultman, D.R.; Gertner, S.N.
 Nature 325, 453-455, 1987
 A/Title: Unusual intron in the immunoglobulin domain of the newly isolated murine CD4 T
 A/Reference number: A26038; MID:87115821; PMID:3027575
 A/Accession: A26038

A: Molecule type: mRNA
 A: Residues: 1-457 <LIT>
 A: Cross-references: GB:X04836; NID:G50353; PIDN:CAA28539.1; PID:G50354
 R: Gorman, S.D.; Tourletille, B.; Parnee, J.R.
 A: Title: Nucleic Acid. Sci. U.S.A. 84, 7644-7648, 1987
 A: Reference number: A39893; MUID:88041159; PMID:2823269
 A: Accession: A39893
 A: Molecule type: DNA
 A: Residues: 1-25, 'E', 27-457 <GOR>
 A: Cross-references: GB:M17080; GB:J03003; NID:G192515; PIDN:AAA37402.1; PID:G387124
 R: Maddon, P.J.; Moliniaux, S.M.; Maddon, D.E.; Zimmermann, K.A.; Godfrey, M.; Alt, F.W.;
 Proc. Natl. Acad. Sci. U.S.A. 84, 9155-9159, 1987
 A: Title: Structure and expression of the mouse gene encoding CD4 and an unusual transcript in brain.
 A: Reference number: A39955; MUID:88097446; PMID:3501122
 A: Accession: A39955
 A: Status: nucleic acid sequence not shown; not compared with conceptual translation
 A: Molecule type: mRNA
 A: Residues: 25-457 <MAD>
 A: Note: the cited GenBank accession number, J03564, is not in release 101.0
 R: Parnee, J.R.; Hunkapiller, T
 Immunol. Rev. 100, 109-127, 1987
 A: Title: L3T4 and the immunoglobulin gene superfamily: New relationships between the imm
 A: Reference number: 154564; MUID:88152875; PMID:3326818
 A: Accession: 154564
 A: Status: translated from GB/EMBL/DBJ
 A: Molecule type: mRNA
 A: Residues: 1-457 <RRS>
 A: Cross-references: GB:M36850; NID:G198670; PIDN:AAA39401.1; PID:G198671
 A: Accession: 169018
 A: Status: translated from GB/EMBL/DBJ
 A: Molecule type: DNA
 A: Residues: 208-318 <RE2>
 A: Cross-references: GB:M36851; NID:G198672; PIDN:AAA39402.1; PID:G554183
 R: Claesson, B.U.; Tsegataros, J.; Kirschbaum, L.; Maddox, J.; Mackay, C.R.; Brandon, M.; M
 Immunogenetics 23, 129-132, 1986
 A: Title: The L3T4 antigen in mouse and the sheep equivalent are immunoglobulin-like.
 A: Reference number: A47642; MUID:86166694; PMID:3082751
 A: Accession: A47642
 A: Molecule type: protein
 A: Residues: 27-43 <CLA>
 C: Comment: This protein is expressed on most thymocytes, on a subset of mature T-cells
 C: Genetics:
 A: Introns: 18/1; 74/1; 128/1; 207/1; 319/1; 386/1; 425/3; 448/2
 C: Superfamily: T-cell surface glycoprotein CD4; Immunoglobulin homology
 C: Keywords: alternative initiator; duplication; glycoprotein; T-cell; transmembrane pro
 F: 1-26/Domain: signal sequence #status predicted <SIG>
 F: 27-457/Product: T-cell surface glycoprotein CD4 #status experimental <MAT>
 F: 35-114/Domain: immunoglobulin homology <IM1>
 F: 119-190/Domain: immunoglobulin homology #status atypical <IM2>
 F: 220-301/Domain: immunoglobulin homology <IM3>
 F: 241-457/Product: CD4, brain-specific short form #status predicted <BRA>
 F: 331-372/Domain: immunoglobulin homology <IM4>
 F: 395-419/Domain: transmembrane #status predicted <TM>
 F: 440-457/Domain: intracellular #status predicted <INT>
 F: 447-112/Domain: intracellular #status predicted <INT>
 F: 167, 288, 323, 392/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 29.1%; Score 993.5; DB 1; Length 457;
 Best Local Similarity 50.9%; Pred. No. 3.6e-51;
 Matches 217; Conservative 66; Mismatches 120; Indels 23; Gaps 8;

QY 1 MNRGVPRRH-LLVVLQALPAAQGNKVLGKGGDTVELTCTASOKKSIOFHKNSNOI 59
 DB 1 MCRARISRRRLLLQLLSQLAVTQKTLVKGESSELPECESSOKKITVFTWKFSROR 60
 QY 60 KILINOG-SFLTGG--PSKINDRADSRSLMDQGNFPIIKNLKIEDSDTYICEVDOKE 116
 DB 61 KILQHGKGVLRIGSGSFG--DRFDSKKGAKWKSFPILIKLMEBSQITICELENRKE 119
 QY 117 EVQLVQLTQSLTLLTQLSGLTLLTLES--PRGSSPSVQCRSPRGKNIQGGKTLVSQLE 175
 DB 120 EELWLVFVKTFSPTGSLILQGGSLTLTLDNSNKSVPNPLECHGKKGKGVVSGSKVLSMNL 179

QY 176 LQDSGTCTVTLQNGKVEFKIDIVLAFQKASIVYKKEGEQVFSFPLAFTEKLTGS 235
 DB 180 VQDSDFNCTVTLQDKQKMGMTLSVGFPGSTATAYKSEGESEFPLNFAE--NGW 237
 QY 236 GELMWQASRASSKSWITFDPLKKNKESVSKVTDOPKQLQMGKPLHLTLPOALPOVAGSG 295
 DB 238 GELMWKAKESKSPQPMWISFSIKKNEVSQKSTDLKQLKETPLPLTKIPQVSLQFAGSG 297
 QY 296 NLTLAEAKTGKHLQEVNLYVNRATQKQKLTCEWGPSPKMLSLKENKAKVSKRE 355
 DB 298 NLTLTLD--KGTILQEVNLYVNRATQKQKLTCEWGPSPKMLSLKENKAKVSKRE 355
 QY 356 KPVVNVNPEAGMOCILSDSGVLTLSNIVLPTWSPVPCAPPEKSCKTHTCPELIG 415
 DB 356 KVVQVNAPEETGLMWQCLLSBGRKAMDRIQVL-----SRVQDTVFLACVLG 402
 QY 416 GPSVFL 421
 DB 403 GSRGFL 408

RESULT 17
 G3MGM
 IG gamma-3 chain C region, membrane-bound form - mouse
 C: Species: Mus musculus (house mouse)
 C: Date: 13-Aug-1986 #sequence_revision 31-Mar-1991 #text_change 16-Jul-1999
 C: Accession: A02156; A02155
 R: Wells, J.A.; Word, C.J.; Rimm, D.; Der-Balan, G.P.; Martinez, H.M.; Tucker, P.W.; Blat
 EMBO J. 3, 2041-2046, 1984
 A: Title: Structure analysis of the murine IgG3 constant region gene.
 A: Reference number: A02156; MUID:85027161; PMID:6092053
 A: Accession: A02156
 A: Molecule type: DNA
 A: Residues: 1-398 <WEL>
 A: Cross-references: GB:J00451; NID:G194392; PIDN:AA859655.1; PID:G194433
 A: Note: the sequence was determined from the germline gene
 R: Komaromy, M.; Claydon, L.; Rogers, J.; Robertson, S.; Kettman, J.; Wall, R.
 Nucleic Acids Res. 11, 6775-6785, 1983
 A: Reference number: A02155; MUID:84041483; PMID:6314258
 A: Accession: A02155
 A: Molecule type: DNA
 A: Residues: 328-332, 'G', 334-341, 'Q', 343-387, 'F', 389-398 <KOM>
 A: Cross-references: GB:K00688
 A: Note: the sequence was determined from the germline gene
 C: Genetics:
 A: Introns: 97/1; 113/1; 223/1; 328/1; 371/3
 C: Complex: An immunoglobulin heterotrimer subunit consists of two identical light (ka
 hain disulfide bonds. In some cases, such as Iga and Igm, the subunits associate into 1
 C: Superfamily: Immunoglobulin C region; immunoglobulin homology
 C: Keywords: alternative splicing; duplication; glycoprotein; heterotrimer; immunoglob
 F: 19-83/Domain: immunoglobulin homology <IM1>
 F: 97-112/Region: hinge
 F: 136-305/Domain: immunoglobulin homology <IM2>
 F: 242-309/Domain: immunoglobulin homology <IM3>
 F: 346-362/Domain: transmembrane #status predicted <TM>
 F: 363-398/Domain: intracellular #status predicted <INT>
 F: 179, 222/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 27.5%; Score 938.5; DB 1; Length 398;
 Best Local Similarity 66.3%; Pred. No. 5.2e-48;
 Matches 169; Conservative 39; Mismatches 40; Indels 7; Gaps 2;

QY 394 VPCGAPRPSGCDTHNCP--ELIGSPVFLFPPPKDMLISRTPEVTCVVNVSDHPE 451
 DB 100 IPRKSTPRGS-----SCPENIILGGSPVFIFFPKPDALMISLTKVTCVVNVSDHPE 154
 QY 452 VKRNVYDGVGVNNAKTKPREBOGNSYTVVSVLTVLHQDMLNGEKYKCVSNKALPAP 511
 DB 155 VHVSWFVDNKEVHTAMQPREADYNSYFRVVSALPIQHQWMGKGEKFCVNNKALPAP 214
 QY 512 EKTISXAKQGRPREQVYTLPRSDDELTKQVSLCLVKGYPEDIAVEMSNQGPENNYK 571

```

Db      215 ERTSKPKGAQCPQVYTTTPRREBMSKKVSLTCLVTNFSALISVENRNELEBDYK 274
      572 TTPPVLDSDGSFPLYSKLTVDKSRMOGVNFCSCVWHEALHNHYTKSLSPGLQDLET 631
      275 NTFPIIDSDGTFLYSKLTVDTSWLGEEIFTCSSVHEALHNHTQKNLSRSPLELNET 334

QY      632 CAEADGDELGLMTT 646
      335 CAEADGDELGLMTT 349

RESULT 18
GIMSW
Ig gamma-1 chain C region, membrane-bound form - mouse
C/Species: Mus musculus (house mouse)
C/Date: 17-Dec-1982 #sequence revision 31-Mar-1991 #text_change 16-Jul-1999
C/Accession: B02159; A02160; B02158
R/Honjo, T.; Obata, M.; Yamawaki-Kataoka, Y.; Kataoka, T.; Kawakami, T.; Takehashi, N.;
Cell 18, 559-568, 1979
A/Title: Cloning and complete nucleotide sequence of mouse immunoglobulin gamma1 chain g
A/Reference number: A02159; MUID:80045036; PMID:115593
A/Accession: B02159
A/Molecule type: DNA
A/Residues: 1-393 <HON>
A/Cross-references: GB:J00453
A/Note: this sequence was determined from the germ-line gene
R/Tyler, B.M.; Cowman, A.F.; Garondakis, S.D.; Adams, J.M.; Bernard, O.
Proc. Natl. Acad. Sci. U.S.A. 79, 2008-2012, 1982
A/Title: mRNA for surface immunoglobulin gamma chains encodes a highly conserved transme
A/Reference number: A02160; MUID:82197626; PMID:6804950
A/Accession: A02160
A/Molecule type: mRNA
A/Residues: 323-393 <TYL>
R/Rogers, J.; Choi, E.; Souza, L.; Carter, C.; Word, C.; Kuehl, M.; Eisenberg, D.; Wall,
Cell 26, 19-27, 1981
A/Title: Gene segments encoding transmembrane carboxyl termini of immunoglobulin gamma C
A/Reference number: A02158; MUID:82115235; PMID:6799207
A/Accession: B02158
A/Molecule type: DNA
A/Residues: 323-366 <ROG>
A/Note: this sequence is the translation of the first exon of the M segment
C/Comment: Cell lines producing IgG contain two mRNA species for Ig gamma chains. The m
nucleotide membrane-bound chains in that it contains an alternative 3' end, encoded in separ
C/Genetics: 1/1; 98/1; 111/1; 218/1; 323/1; 366/3
A/Functions: 1/1; 98/1; 111/1; 218/1; 323/1; 366/3
C/Complex: An immunoglobulin heterotrimer subunit consists of two identical light (kappa)
hain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into 1a
C/Superfamily: immunoglobulin C region; immunoglobulin homology
C/Keywords: alternative splicing; duplication; glycoprotein; heterotrimer; immunoglobuli
F/331-200/Domain: immunoglobulin homology <IMM>
F/340-357/Domain: transmembrane #status predicted <TM>
F/358-393/Domain: intracellular #status predicted <INT>
F/174,278/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match      27.2%; Score 928.5; DB 1; Length 393;
Best Local Similarity 57.2%; Pred. No. 2e-47;
Matches 167; Conservative 51; Mismatches 47; Indels 27; Gaps 3;

QY      379 LLESNIKVLPTWSTPVPCCAP-EPKSCDKTH-----TCPELL 414
      56 VLQSDLYTLSSSVTPSSPRSETVTCNVNHPASTKVDKIVPRDGCPCICTYPEV 114
      415 GGPSTVLPFPKPYDTLMTISTPEVTCVVDVSHEDPEVKNNYVGVENAKTKRREO 474
      115 --SSVFIFPKPKPDVLTITLPKTCVVDVSKDDEVPQSWFVDVVEVTAQTORREO 172
      475 VNSTVAVSVLTVLHDMNGKVKCKVSNKALPAPLEKTIKAKGQRPQVYTLPPSR 534
      173 FNSFSSVSELPIMHQMNGKFKCKVNSAALPAPLEKTIKTKGPKAPQVYTLPPK 232
      535 DELTKQVSLTCLVKGFPSPDIAVENWESNGQFENNYKTPPVLDSDGSFPLYSKLTVDKS 594
      ::: ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::

```

```

Db      233 EGMAKQVSLTGMITDFPEFDITVEWQNGQPAENYKNTQPIKNTNGSYFVYSKLVOKS 292
      595 RMQGVNFCSCVWHEALHNHYTKSLSPGLQDLETCAEADGDELGLMTT 646
      293 NMEAGNTFTCSVLHRLGNHHTKSLSPGLQDLETCAEADGDELGLMTT 344

QY      293 NMEAGNTFTCSVLHRLGNHHTKSLSPGLQDLETCAEADGDELGLMTT 344

RESULT 19
GHRB
Ig gamma chain C region - rabbit
C/Species: Oryctolagus cuniculus (domestic rabbit)
C/Date: 24-Apr-1984 #sequence revision 15-Nov-1984 #text_change 16-Jul-1999
C/Accession: A91749; A90290; A93928; A90245; A94416; A02161
R/Bernstein, K.E.; Alexander, C.B.; Mage, R.G.
Immunogenetics 18, 387-397, 1983
A/Title: Nucleotide sequence of a rabbit IgG heavy chain from the recombinant F-1 haplo-
A/Reference number: A91749; MUID:84030930; PMID:6313520
A/Accession: A91749
A/Molecule type: mRNA
A/Residues: 1-333 <BER>
A/Note: this sequence has the d12 allotypic marker, 104-Thr, and the e14 marker, 185-Th.
R/Pratt, D.M.; Mole, L.E.
Biochem. J. 151, 337-349, 1975
A/Title: Sequence studies on the constant region of the Fd sections of rabbit immunoglob
A/Reference number: A90290; MUID:76135469; PMID:1243651
A/Accession: A90290
A/Molecule type: protein
A/Residues: 1-47; 'E', 49-71, 'PV', 72-128 <PRA>
R/Wartens, C.L.; Moore, K.W.; Steinmetz, M.; Hood, L.; Knight, K.L.
Proc. Natl. Acad. Sci. U.S.A. 79, 6018-6022, 1982
A/Title: Heavy chain genes of rabbit IgG; isolation of a cDNA encoding gamma heavy chain
A/Reference number: A93928; MUID:83299917; PMID:6193512
A/Accession: A93928
A/Molecule type: mRNA
A/Residues: 88-103; 'M', 105-143, 'E', 145-184, 'A', 186, 'E', 188-266 <MAR>
A/Cross-references: GB:M16426; MUID:9165111; PIDN:AA31289.1; PID:9165112
A/Note: this sequence has the d11 allotypic marker, 104-Met, and the e15 allotypic mark
R/Pruchter, R.G.; Jackson, S.A.; Mole, L.E.; Porter, R.R.
Biochem. J. 116, 249-259, 1970
A/Title: Sequence studies of the Fd section of the heavy chain of rabbit immunoglobulin
A/Reference number: A90245; MUID:70110015; PMID:5461106
A/Accession: A90245
A/Molecule type: protein
A/Residues: 132-143; 'E', 145-161 <FRU>
R/Hill, R.L.; Lebovitz, H.E.; Fellows Jr., R.E.; Delaney, R.
In Gamma Globulins, Nobel Symp. 3, Killander, J., ed., pp.109-127, Almqvist and Wiksell,
A/Reference number: A94416
A/Accession: A94416
A/Molecule type: protein
A/Residues: 129-131; 135-172, 'D', 174-184, 'A', 186, 'E', 188-200, 'D', 202-217, 'E', 219-232, 'Q',
A/Note: this has the e15 allotypic marker, 185-Ala
C/Complex: An immunoglobulin heterotrimer subunit consists of two identical light (kappa)
hain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into 1
C/Superfamily: immunoglobulin C region; immunoglobulin homology
C/Keywords: duplication; glycoprotein; heterotrimer; immunoglobulin
F/20-82/Domain: immunoglobulin homology <IM1>
F/130-199/Domain: immunoglobulin homology <IM2>
F/336-303/Domain: immunoglobulin homology <IM3>
F/173/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match      26.7%; Score 910; DB 1; Length 323;
Best Local Similarity 65.7%; Pred. No. 1.9e-46;
Matches 174; Conservative 31; Mismatches 44; Indels 16; Gaps 5;

QY      375 SGQVLLSNIKVLPTWSTPVPCC-----PKSCDKTHTC--PELLGGPSVF 420
      60 SGYLSLSVSVTS--SGPVCNVAPATNTKVDKTVASTCSKP--TCPELLGGPSVF 117
      421 LFPKPKDTLMTSRPEVTCVVDVSHEDPEVKNNYVGVENAKTKRREOYSTVR 480
      118 LFPKPKDTLMTSRPEVTCVVDVSDDEVPQFTVYINNEOVRTARPLREOQFNSTIR 177
      481 VVSVLTVLHDMNGKVKCKVSNKALPAPLEKTIKAKGQRPQVYTLPPSRDELTKN 540
      ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::

```

```

Db 178 VVSTLPTIHQDWLKRKEKCKVHNKALPAPIEKTISKARGPLBPKVYTMGPPEEELSSR 237
OY 541 QVSTLCLVKGFPYSDIAVEMESGPENNYKTPTPVLDSDGSFPLYSKLTVDKSRMOGN 600
Db 238 SVSLTCLMNGYFPDSIDISEWENKGAEDNYKTPPAVLDSDSYFLNKLSPVISEMORG 297
OY 601 VFSCSVMEHALNHNHYTKSLSPG 625
Db 298 VFTCSVMHEALHNHYTKSLSPG 322

```

RESULT 20

```

G2MSAM
Ig gamma-2a chain C region, membrane-bound form - mouse
C:Species: Mus musculus (house mouse)
C:Date: 19-Feb-1994 #sequence revision 31-Mar-1991 #text_change 16-Jul-1999
C:Accession: A02154; B32657; I57809
R:Yamawaki-Kataoka, Y.; Nakai, S.; Miyata, T.; Honjo, T.
Proc. Natl. Acad. Sci. U.S.A. 79, 2623-2627, 1982
A:Title: Nucleotide sequences of gene segments encoding membrane domains of immunoglobulin
A:Reference number: A02154; MUID:82222190; PMID:6283537
A:Accession: A02154
A:Molecule type: DNA
A:Residues: 329-399 <YAM>
A:Cross-references: GB:J00471
A:Note: the sequence was determined from the germ-line gene
R:Yamawaki-Kataoka, Y.; Miyata, T.; Honjo, T.
Nucleic Acids Res. 9, 1365-1381, 1981
A:Title: The complete nucleotide sequence of mouse immunoglobulin gamma 2 A gene and env
A:Reference number: A32657; MUID:81198976; PMID:6262729
A:Accession: B32657
A:Molecule type: DNA
A:Residues: 1-329, 'K' <YAM>
R:Hall, B.; Milcarek, C.
Mol. Immunol. 26, 819-826, 1989
A:Title: Sequence and polyadenylation site determination of the murine immunoglobulin gamma
A:Reference number: I57809; MUID:9007953; PMID:2513486
A:Accession: I57809
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 373-399 <RES>
A:Cross-references: GB:M5032; NID:9194478; PIDN:AAA37919.1; PID:9387217
A:Comment: The sequence of residues 1-328 was assumed to be identical with the counterpart
C:Comment: Cell lines producing IgG contain two mRNA species for Ig gamma chains. The major
C:Comment: chain contains an alternative 3' end, encoded in separate exons, that is homologous with
C:Genetics:
A:Introns: 1/1; 98/1; 114/1; 224/1; 329/1; 372/1
C:Complex: An immunoglobulin heterotrimer subunit consists of two identical light (kap)
C:Superfamily: immunoglobulin C region; immunoglobulin homology
C:Keywords: alternative splicing; duplication; glycoprotein; heterotrimer; immunoglobulin
F:137-206/domain: immunoglobulin homology <IMM>
F:346-353/domain: transmembrane #status predicted <TM>
F:354-359/domain: intracellular #status predicted <INT>
F:180/Binding site: carbohydrate (asn) (covalent) #status predicted

```

```

Query Match 26.4%; Score 902.5; DB 1; Length 399;
Best Local Similarity 49.0%; Pred. No. 6.9e-46;
Matches 180; Conservative 45; Mismatches 87; Indels 55; Gaps 5;

```

```

OY 285 PQALPOYAGSNLTALAEKTKLHGEV-----NLVWRAKATOLQKLTCEVWGTSTKLM 339
Db 34 PEPPTLTWNSGSLSSGHTPPAVLQSDLYLTSSSVTYSSTWPSQITCNVAHPAS---- 89
OY 340 LSLKLENKAKVSKREKPVWVNLNPEAGMOCCLSDSQVLLSNIKYLPTWSTPVPAP 399
Db 90 -----STKVDK-----KLEPQPTIKPCP-- 108
OY 400 EPKSCDKTHTCPELLGSPVFLFPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNYYVD 459
Db 109 -PKCK-----PAPNLTGSPVIFPPKIKDVLMISSLPSIVTCVVVDVSEDDPDVQISWFIN 163

```

```

OY 460 GVEVHNKATKPREQNVSTYRVVSVLTTLHQMINKKEYKCVSNKALPAPIEKTISKAK 519
Db 164 NVEVHTAQTOHREDYVNSTLRVVSALPIHQDMWSGKEFKCVKNKOLPAPIERTISKPK 223
OY 520 GQREPEPVYTLPPSRDRLTKQVSLTCLVYGFPSDIAVEMESGPENNYKTPTPVLD 579
Db 224 GSVAPQVYVLPPEEEMTKQVTLTCMVDFMEDLYVMTNNGKTELNYKTEPVLDS 283
OY 580 DGSFFLYSKLTVDKSRMOGNVSCSVMEHALNHNHYTKSLSPGLQDETCAEAQDGE 639
Db 284 DGSYFMYSKLRVEKQKQWENRNSYSCSVMEHALNHNHYTKSLSPGLQDETCAEAQDGE 343
OY 640 LDGLMTT 646
Db 344 LDGLMTT 350

```

RESULT 21

```

147160
Ig gamma 2b chain constant region - pig (fragment)
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 21-Feb-1997 #sequence revision 21-Feb-1997 #text_change 21-Jan-2000
C:Accession: I47160
R:Kachkovics, I.; Sun, J.; Butler, J.E.
J. Immunol. 153, 3565-3573, 1994
A:Title: Five putative subclasses of swine IgG identified from the cDNA sequences of a
A:Reference number: I47158; MUID:95015845; PMID:7930579
A:Accession: I47160
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-328 <KAC>
A:Cross-references: EMBL:U03780; NID:9433125; PIDN:AA52218.1; PID:9433126
C:Genetics:
A:Gene: IgG2b
C:Superfamily: immunoglobulin C region; immunoglobulin homology
F:133-202/Domain: immunoglobulin homology <IMM>

```

```

Query Match 26.4%; Score 901.5; DB 2; Length 328;
Best Local Similarity 50.1%; Pred. No. 6.2e-46;
Matches 185; Conservative 46; Mismatches 73; Indels 65; Gaps 8;

```

```

OY 265 RVTQDPQLQMGKLPHTLTPQALPOYAG-----SGNLTALAEKTKLHGEVNLVWRAKATQ 321
Db 16 RDTSGPVALGCLASSYFPPEPVYTNWNSGALTYGHTFPVLDPSGLYSLSSTVTPASS 75
OY 322 L-QKLTCEVWG-TPSKMLSLKENKAKVSKREKPVWVNLNPEAGMOCCLSDSQVL 379
Db 76 LSKSYTCNVNHPATTKV-----DKRVGKTKP----- 104
OY 380 LESNIKLPTWSTPVPAPAPKSCDKTHTCPELLG-GPSVFLFPKPKDTLMISRTPEV 438
Db 105 -----PCP-----ICPACESGPEVFFPPPKDTLMISRTPEV 138
OY 439 TCVVVDVSHEDPEVKFNWYVDGVEVHNKATKPREEOVNSTYRVVSVLTTLHQMINKKEY 498
Db 139 TCVVVDVSGNPEVQSVYDGVHTAQKPREEOVNSTYRVVSVLTPIHQDMINKKEF 198
OY 499 KCVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAV 558
Db 199 KCVNNKNDAPAPIETIRISKAKGQPREPOVYTLPPHAEELSRKVSITCLVIGYPPDIV 258
OY 559 EMESSNGO--PENNYKTPTPVLDSDGSFPLYSKLTVDKSRMOGNVSCSVMEHALNHNHY 616
Db 259 EWRQNGQPEDEGVRITTPQQDVDTGYFLYSKPSVDKASWGGGIFQCAVMHEALHNHYT 318
OY 617 QKSLSLSPG 625
Db 319 QKSLSLTPG 327

```

```

RESULT 22
147159
Ig gamma 2a chain constant region - pig (fragment)

```


A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-472 <PART>
 A:Cross-references: EMBL:X69797
 C:Superfamily: immunoglobulin C region; immunoglobulin homology
 C:Keywords: immunoglobulin
 F:277-346/Domain: immunoglobulin homology <IMM>

Query Match 24.7%; Score 843; DB 2; Length 472;
 Best Local Similarity 41.2%; Pred. No. 2.7e-42;
 Matches 213; Conservative 59; Mismatches 141; Indels 104; Gaps 17;

QY 157 SPKGNIGGKTLVSQLELDSGTWCTVLQNGKV-----EFKIDIVLAFQKA 207
 DB 11 APRG-----VLGQVRLQESGSPSLATLLQTLSTCTISGFSLNNGVDMVWQARCKA 61
 QY 208 SSIYKKEGEVERSFPLATVEKLTGSG---ELMQAERASSKSNITPDLKKEVSXK 264
 DB 62 -----LEWLGSGYDEDIDYND-VLKRSLITTKDTSKQVSLT 98
 QY 265 R---VTQDPKQLQMGKCLPLHLTLPLQALPOVA--GSGNLTLLBAKTKLHQEVNLVYMR 319
 DB 99 LSTVTEDTAVYCARVDYDSSHAFAASYPKPGGLISVLSAST-----TP 146
 QY 320 TQLOKNLTCVWGPTSPKMLSLKENKAKYKREKRVWY-----LNPBAGMOCLLS 373
 DB 147 PKVYPLTSC--CGDTSSSIVTLGCLVS---SYMPEPVTVTNMAGALTSGVHTFPALQ 199
 QY 374 DSGVLLSNNIKV-----LPTWSTPVP-----CPAPEKSCDKTHTC-- 410
 DB 200 SSGYSLSSSVTVPASTGAGTFCINVAHPASSTKVRVPGGPDPC-KKCC---RCPP 254
 QY 411 PELGSGPSVFLFPKPKDQTLMSRTPEYTCVNVVSHEDPEKFNWYVDGEVNAKTKP 470
 DB 255 PELGSGPSVFLFPKPKDQTLMSRTPEYTCVNVVSHEDPEKFNWYVDGEVNAKTKP 314
 QY 471 REEQYNSTYRVSVLTVLHODMLNGEKYCKVSNKALPAPIEKTISAKGQPREPOVYTL 530
 DB 315 REEQFNSTFRVVSALPIQHOMWGKEPKCKVHNHALPAPIVRTISTKQARERPOVYVL 374
 QY 531 PRSDELTKNOVSLTCLVKGFPESDIAMVESNQP--ENNYKTTTPYLDSDGSFFLYSK 588
 DB 375 APPEBELSKSTLSTYCLVTGFPYDIAVEMQKQPESEDEYGTTSQDLADSGSYFYSR 434
 QY 589 LTVDKSRMOGNYFSCSVMEALHNHYTKSLSPG 625
 DB 435 LRVDKSMQEGDTYACVMEALHNHYTKSLSPG 471

RESULT 30

G3MGC
 Ig gamma-3 chain C region, secreted form - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 17-Mar-1987 #sequence_revision 31-Mar-1991 #text_change 16-Jul-1999
 C:Accession: B02156
 R:Mele, J.A.; Word, C.J.; Rimm, D.; Der-Balan, G.P.; Martinez, H.M.; Tucker, P.W.; Blact
 EMBL J. 3, 2041-2046, 1984
 A:Title: Structure analysis of the murine IgG3 constant region gene.
 A:Reference number: A02156; MUID:85027161; PMID:6092053
 A:Molecule type: DNA
 A:Residues: 1-329 <WEL>
 A:Cross-references: GB:J00451
 A>Note: the sequence was determined from the germline gene
 C:Genetics:
 A:Insertions: 97/1; 113/1; 223/1
 C:Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kap
 hain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into 1a
 C:Superfamily: immunoglobulin C region; immunoglobulin homology
 C:Keywords: alternative splicing; duplication; glycoprotein; heterotetramer; immunoglobul
 F:19-83/Domain: immunoglobulin homology <IM1>
 F:97-112/Region: hinge
 F:136-205/Domain: immunoglobulin homology <IM2>

F:242-309/Domain: immunoglobulin homology <IM3>
 F:179,322/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 24.6%; Score 840.5; DB 1; Length 329;
 Best Local Similarity 64.5%; Pred. No. 2.4e-42;
 Matches 151; Conservative 37; Mismatches 39; Indels 7; Gaps 2;
 QY 394 VPCAPPEKSCDKTHTC--ELGSPSVFLFPKPKDQTLMSRTPEYTCVNVVSHEDPE 451
 DB 100 IPRSTPFGS-----SCPBNITLGGPSVFLFPKPKDQTLMSRTPEYTCVNVVSHEDPD 154
 QY 452 VKENWYDGEVHNATKPREEQYNSTYRVSVLTVLHODMLNGEKYCKVSNKALPAPI 511
 DB 155 VHSWVFVNDKEVHTAWTQPREAQYNSTFRVVSALPIQHOMWGKEPKCKVHNHALPAPI 214
 QY 512 EKTISAKGQPREPOVYTLPRSDELTKNOVSLTCLVKGFPESDIAMVESNQPENNYK 571
 DB 215 ERTISKPKGAAQTPQVTVTTPPRBQMSKKVSLCLVTNPFSEATISVERNGBLEDQYK 274
 QY 572 TTPPYLDSDGSFFLYSKLTVDKSRMOGNYFSCSVMEALHNHYTKSLSPG 625
 DB 275 NTPPIIDSDGTIFLYSKLTVDTDSMGLGELFTCSVMEALHNHYTKSLSPG 328

RESULT 31

C30554
 Ig heavy chain C region - sheep (fragment)
 C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
 C:Date: 03-Mar-1989 #sequence_revision 03-Mar-1989 #text_change 21-Jan-2000
 C:Accession: C30554
 R:Foley, R.C.; Beh, K.J.
 J. Immunol. 142, 708-711, 1989
 A:Title: Isolation and sequence of sheep Ig H and L chain cDNA.
 A:Reference number: A30554; MUID:89093962; PMID:2492052
 A:Accession: C30554
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: mRNA
 A:Superfamily: immunoglobulin C region; immunoglobulin homology
 C:Keywords: heterotetramer; immunoglobulin
 F:113-182/Domain: immunoglobulin homology <IM>

Query Match 24.5%; Score 836; DB 2; Length 308;
 Best Local Similarity 49.1%; Pred. No. 4e-42;
 Matches 173; Conservative 39; Mismatches 76; Indels 64; Gaps 7;

QY 284 LPQALPOYAGSGLTL-----ALEAKTKLHQEVNLVYMRATQLOKNLTCVWGPTSPK 337
 DB 10 MPEPVTVTNMAGALTSGVHTFPALQSSGLYSVTVPASTGAGTFCINVAHPAS-- 67
 QY 338 LMLSLKENKAKYKREKRVWVLPBAGMOCLSDSGVLLSNNIKVLPWSTPVP 397
 DB 68 -----STVYDKRVBP-----GCPDPC- 83

QY 398 APPEKSCDKTHTC--PELGGPSVFLFPKPKDQTLMSRTPEYTCVNVVSHEDPEKFN 455
 DB 84 -----KPC-----RCPPELPGGSPVFLFPKPKDQTLMSRTPEYTCVNVVSHEDPEKFN 135
 QY 456 WYVDGEVHNATKPREEQYNSTYRVSVLTVLHODMLNGEKYCKVSNKALPAPIEKT 515
 DB 136 WYVDGEVHNATKPREEQYNSTYRVSVLTVLHODMLNGEKYCKVSNKALPAPIEKT 195
 QY 516 SKAGQPREPOVYTLPRSDELTKNOVSLTCLVKGFPESDIAMVESNQP--ENNYKTT 573
 DB 196 SRTKGAAREPOVYTLPRSEBELTKNOVSLTCLVKGFPESDIAMVESNQP--ENNYKTT 255
 QY 574 PLYLDSDGSFFLYSKLTVDKSRMOGNYFSCSVMEALHNHYTKSLSPG 625
 DB 256 TSQDLADSGFFLYSKLTVDKSRMOGNYFSCSVMEALHNHYTKSLSPG 307

RESULT 32

PC4436

A>Title: Evolution of immunoglobulin subclasses. Primary structure of a murine myeloma g
A/Reference number: A26237; MUID:78242288; PMID:98524
A/Contents: annotation; WOPC 21
A/Note: this is the final paper in a series reporting the protein sequence, the disulfid
A/Note: there are a number of differences from the sequence shown
C/Genetics:
A/Introns: 1/1; 98/1; 111/1; 218/1
C/Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kap
hain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into 1a
C/Superfamily: Immunoglobulin C region; immunoglobulin homology
C/Keywords: alternative splicing; duplication; glycoprotein; heterotetramer; immunoglob
F/20-84/Domain: immunoglobulin homology <IM1>
F/98-110/Region: hinge
F/131-200/Domain: immunoglobulin homology <IM2>
F/237-304/Domain: immunoglobulin homology <IM3>
F/27-82,138-198,244-302/Disulfide bonds: #status experimental
F/102/Disulfide bonds: interchain (to light chain) #status experimental
F/104,107,109/Disulfide bonds: interchain (to heavy chain) #status experimental
F/174/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 23.9%; Score 814.5; DB 1; Length 324;
Best Local Similarity 53.9%; Pred. No. 7.9e-41;
Matches 146; Conservative 51; Mismatches 47; Indels 27; Gaps 3;

Qy 379 LIESNKKVLTWSTPRPCAP-EPKSGDKTH-----TCPELL 414
Db 56 VLQSDLYTSSSVTVSSPSPSTVTCVNAHPASSTKVDKIVPDCCKPCCTVPEV- 114
Qy 415 GGSFVLFPPKPKDTLMISRTPEVTCVVDVSHEDPEVKFNWYDGEVNAKTPREQ 474
Db 115 --SSVFPPKPKPDVLTITLPKVTCTVVDISKDEPEVQSWVDVENVTAQTQPREQ 172
Qy 475 YNSTYRVSVLTVLHODMLNGKEYKCKVSKALPAPLEKTTSAKQPREPVYTLPPSR 534
Db 173 FNSTFVSSELPTMHODMLNGKEFKCKVNSAAPAPLEKTTSAKQPREPVYTLPPSR 232
Qy 535 DELTKNOVSLTCLVKGYRPSDIVEMESNGCPENNYKTPPVLDGSGFPLYSKLVDS 594
Db 233 EQPAKDVSLTCTITPFPEDITVEWQNPAPENNYKNTPIWNTGSIFFVSKLVQKS 292
Qy 595 RMOGNVFCGVNAHEALHNHYTKSLSPG 625
Db 293 NWEAGNFTCSVLHEGHNNHTKSLSHSG 323

RESULT 38
G2MS11
Ig gamma-2b chain - mouse
C/Species: Mus musculus (house mouse)
C/Date: 31-Mar-1980 #sequence, revision 01-Dec-2000 #text change 01-Dec-2000
C/Accession: S25057; A02157; A26232; A26233; A53598
R/Fischer, R.; Voss, A.; Nierbach, M.; Munziker, W.; Hirsch, H.J.; Kreuzaler, F.
submitted to the EMBL Data Library, July 1992
A/Description: Production of a Tobacco mosaic virus (TMV) inactivating neotop specific n
A/Reference number: S25057
A/Accession: S25057
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-474 <FIS>
R/Yamawaki-Kataoka, Y.; Kataoka, T.; Takahashi, N.; Obata, M.; Honjo, T.
Nature 283, 786-789, 1980
A/Title: Complete nucleotide sequence of immunoglobulin gamma2b chain gene cloned from m
A/Reference number: A02157; MUID:80120716; PMID:6766534
A/Contents: a allele
A/Accession: A02157
A/Molecule type: DNA
A/Residues: 138-161, 'L', 163-189, 'FP', 193-474 <YAM>
A/Cross-references: GB:J00461
A/Note: the sequence was determined from the germ-line gene
R/Tucker, P.W.; Marcu, K.B.; Slightom, J.L.; Blattner, F.R.
Science 206, 1299-1303, 1979
A/Title: Structure of the constant and 3' untranslated regions of the murine gamma2b hea

A/Reference number: A26235; MUID:80081501; PMID:117548
A/Contents: MPC 11
A/Accession: A26235
A/Molecule type: mRNA
A/Residues: 138-172, 'P', 174-189, 'FP', 193-376, 'T', 378-474 <TU>
A/Note: Lys-474 is probably removed posttranslationally
R/Tucker, P.W.; Marcu, K.B.; Newell, N.; Richards, J.; Blattner, F.R.
Science 206, 1303-1306, 1979
A/Title: Sequence of the cloned gene for the constant region of murine gamma2b immunogl
A/Reference number: A26232; MUID:80081502; PMID:117549
A/Accession: A26232
A/Molecule type: DNA
A/Residues: 138-172, 'P', 174-189, 'FP', 193-376, 'T', 378-474 <TU>
R/Ollo, R.; Rougeon, F.
Nature 286, 761-763, 1982
A/Title: Mouse immunoglobulin allotypes: post-duplication divergence of gamma2a and gam
A/Reference number: A26233; MUID:82173203; PMID:6803173
A/Contents: b allele
A/Accession: A26233
A/Molecule type: DNA
A/Residues: 138-161, 'L', 163-189, 'FP', 193-300, 'R', 302-331, 'N', 333-437, 'DI', 440-474 <OLL>
A/Cross-references: GB:J00461
R/Kim, H.; Yamaguchi, Y.; Maeda, K.; Matsunaga, C.; Yamamoto, K.; Irimura, T.; Takahashi
J. Biol. Chem. 269, 12345-12350, 1994
A/Title: O-glycosylation in hinge region of mouse immunoglobulin G2b.
A/Reference number: A53598; MUID:94216359; PMID:7512967
A/Accession: A53598
A/Status: preliminary
A/Molecule type: protein
A/Residues: 234-251 <KIM>
C/Comment: The a allele sequence is shown.
C/Genetics:
A/Introns: 138/1; 236/1; 258/1; 368/1
C/Complex: An immunoglobulin heterotetramer subunit consists of two identical light (ka
hain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into 1a
C/Superfamily: Immunoglobulin C region; immunoglobulin homology
C/Keywords: alternative splicing; duplication; glycoprotein; heterotetramer; immunoglob
F/157-222/Domain: immunoglobulin homology <IM1>
F/236-257/Region: hinge
F/281-350/Domain: immunoglobulin homology <IM2>
F/387-454/Domain: immunoglobulin homology <IM3>
F/152/Disulfide bonds: interchain (to light chain) #status predicted
F/164-220,288-348,394-452/Disulfide bonds: #status predicted
F/247,250,255,256/Disulfide bonds: interchain (to heavy chain) #status predicted
F/324/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 23.6%; Score 807; DB 1; Length 474;
Best Local Similarity 33.9%; Pred. No. 3.6e-40;
Matches 205; Conservative 73; Mismatches 149; Indels 178; Gaps 17;

Qy 34 GDTVELTCTASQKSIQF--HKNSNOIKILNQG---SFL--TKGPKLMDRADSRSS 85
Db 34 GASVKNSCASGYTFTTYVMHW-----VKQKPGGIEWIGYINPNKDGTFNEKFKGKAT 88
Qy 86 LW---DQGNFPLINKLKTEDSDTYICEVEDQKEVQVLLFGLTANSPTNHLQGGSLTLT 142
Db 89 LTSKSSNTAYWELSLTSEDSAYVYCARD-----YDYDFATWGGSTVLT 134
Qy 143 LESPSSPSVOCSPRGKNIQGGKTLVSQLELDSGTWCTVLONQKVEFKIDIVVL 202
Db 135 VSAAKTTPSVYPLAPGCGDITGSSVTS----- 162
Qy 203 APOKASSTIYKKEGEVRSFPLATVEKLTGSGELMQAEPASSSKSTITDLKKEVS 262
Db 163 -----GCLVKKYPPESVTVT-----NNSGSLSS- 186
Qy 263 VKRVTQDPKLGQKKLPHLTLPOLPOVAGSNLTALBAKTKGKHQEVNLVVMKATOL 322
Db 187 -----VH-TLSQALQ---SGLYTMS-----SSVTPGSTVP 214
Qy 323 OKNLTCGVNTPSKMLSLKLENKAKYSKREKRVVNLNPAQMWQCULSDSGQVLLS 382
Db 215 SQTVTCSVNAHPAS-----STVYDKLEP-----SGPI----- 241